Linear Algebra



David Cherney, Tom Denton,
Rohit Thomas and Andrew Waldron

Edited by Katrina Glaeser and Travis Scrimshaw
First Edition. Davis California, 2013.

Contents

1

What is Linear Algebra?

Many difficult problems can be handled easily once relevant information is organized in a certain way. This text aims to teach you how to organize information in cases where certain mathematical structures are present. Linear algebra is, in general, the study of those structures. Namely

> Linear algebra is the study of vectors and linear functions.

In broad terms, vectors are things you can add and linear functions are functions of vectors that respect vector addition. The goal of this text is to teach you to organize information about vector spaces in a way that makes problems involving linear functions of many variables easy. (Or at least tractable.)

To get a feel for the general idea of organizing information, of vectors, and of linear functions this chapter has brief sections on each. We start here in hopes of putting students in the right mindset for the odyssey that follows; the latter chapters cover the same material at a slower pace. Please be prepared to change the way you think about some familiar mathematical objects and keep a pencil and piece of paper handy!

1.1 Organizing Information

Functions of several variables are often presented in one line such as

$$f(x,y) = 3x + 5y\,.$$

But lets think carefully; what is the left hand side of this equation doing? Functions and equations are different mathematical objects so why is the equal sign necessary?



A Sophisticated Review of Functions

If someone says

"Consider the function of two variables $7\beta - 13b$."

we do not quite have all the information we need to determine the relationship between inputs and outputs.

Example 1 (Of organizing and reorganizing information)
You own stock in 3 companies: $Google$, $Netflix$, and $Apple$. The value V of your stock portfolio as a function of the number of shares you own s_N, s_G, s_A of these companies is

$$24s_G + 80s_A + 35s_N\,.$$

Here is an ill posed question: what is $V\begin{pmatrix}1\\2\\3\end{pmatrix}$?

The column of three numbers is ambiguous! Is it is meant to denote

- 1 share of G, 2 shares of N and 3 shares of A?
- 1 share of N, 2 shares of G and 3 shares of A?

Do we multiply the first number of the input by 24 or by 35? No one has specified an order for the variables, so we do not know how to calculate an output associated with a particular input.[1]

A different notation for V can clear this up; we can denote V itself as an ordered triple of numbers that reminds us what to do to each number from the input.

[1] Of course we would know how to calculate an output if the input is described in the tedious form such as "1 share of G, 2 shares of N and 3 shares of A", but that is unacceptably tedious! We want to use ordered triples of numbers to concisely describe inputs.

Denote V by $\begin{pmatrix}24 & 80 & 35\end{pmatrix}$ and thus write $V\begin{pmatrix}1\\2\\3\end{pmatrix}_B = \begin{pmatrix}24 & 80 & 35\end{pmatrix}\begin{pmatrix}1\\2\\3\end{pmatrix}$

to remind us to calculate $24(1) + 80(2) + 35(3) = 334$

because we chose the order $\begin{pmatrix}G & A & N\end{pmatrix}$ and named that order B

so that inputs are interpreted as $\begin{pmatrix}s_G\\s_A\\s_N\end{pmatrix}$.

If we change the order for the variables we should change the notation for V.

Denote V by $\begin{pmatrix}35 & 80 & 24\end{pmatrix}$ and thus write $V\begin{pmatrix}1\\2\\3\end{pmatrix}_{B'} = \begin{pmatrix}35 & 80 & 24\end{pmatrix}\begin{pmatrix}1\\2\\3\end{pmatrix}$

to remind us to calculate $35(1) + 80(2) + 24(3) = 264$.

because we chose the order $\begin{pmatrix}N & A & G\end{pmatrix}$ and named that order B'

so that inputs are interpreted as $\begin{pmatrix}s_N\\s_A\\s_G\end{pmatrix}$.

The subscripts B and B' on the columns of numbers are just symbols[2] reminding us of how to interpret the column of numbers. But the distinction is critical; as shown above V assigns completely different numbers to the same columns of numbers with different subscripts.

There are six different ways to order the three companies. Each way will give different notation for the same function V, and a different way of assigning numbers to columns of three numbers. Thus, it is critical to make clear which ordering is used if the reader is to understand what is written. Doing so is a way of organizing information.

[2]We were free to choose any symbol to denote these orders. We chose B and B' because we are hinting at a central idea in the course: choosing a basis.

This example is a hint at a much bigger idea central to the text; our choice of order is an example of choosing a *basis*[3].

The main lesson of an introductory linear algebra course is this: you have considerable freedom in how you organize information about certain functions, and you can use that freedom to

1. uncover aspects of functions that don't change with the choice (Ch 12)
2. make calculations maximally easy (Ch 13 and Ch 17)
3. approximate functions of several variables (Ch 17).

Unfortunately, because the subject (at least for those learning it) requires seemingly arcane and tedious computations involving large arrays of numbers known as matrices, the key concepts and the wide applicability of linear algebra are easily missed. So we reiterate,

> Linear algebra is the study of vectors and linear functions.

In broad terms, vectors are things you can add and linear functions are functions of vectors that respect vector addition.

1.2 What are Vectors?

Here are some examples of things that can be added:

Example 2 (Vector Addition)

(A) Numbers: Both 3 and 5 are numbers and so is $3+5$.

(B) 3-vectors: $\begin{pmatrix}1\\1\\0\end{pmatrix}+\begin{pmatrix}0\\1\\1\end{pmatrix}=\begin{pmatrix}1\\2\\1\end{pmatrix}$.

[3] Please note that this is an *example* of choosing a basis, not a statement of the definition of the technical term "basis". You can no more learn the definition of "basis" from this example than learn the definition of "bird" by seeing a penguin.

(C) Polynomials: If $p(x) = 1 + x - 2x^2 + 3x^3$ and $q(x) = x + 3x^2 - 3x^3 + x^4$ then their sum $p(x) + q(x)$ is the new polynomial $1 + 2x + x^2 + x^4$.

(D) Power series: If $f(x) = 1+x+\frac{1}{2!}x^2+\frac{1}{3!}x^3+\cdots$ and $g(x) = 1-x+\frac{1}{2!}x^2-\frac{1}{3!}x^3+\cdots$ then $f(x) + g(x) = 1 + \frac{1}{2!}x^2 + \frac{1}{4!}x^4 \cdots$ is also a power series.

(E) Functions: If $f(x) = e^x$ and $g(x) = e^{-x}$ then their sum $f(x) + g(x)$ is the new function $2\cosh x$.

There are clearly different kinds of vectors. Stacks of numbers are not the only things that are vectors, as examples C, D, and E show. Vectors of different kinds can not be added; What possible meaning could the following have?

$$\begin{pmatrix} 9 \\ 3 \end{pmatrix} + e^x$$

In fact, you should think of all five kinds of vectors above as different kinds, and that you should not add vectors that are not of the same kind. On the other hand, any two things of the same kind "can be added". This is the reason you should now start thinking of all the above objects as vectors!

In Chapter 5 we will give the precise rules that vector addition must obey. In the above examples, however, notice that the vector addition rule stems from the rules for adding numbers.

When adding the same vector over and over, for example

$$x + x\,,\quad x + x + x\,,\quad x + x + x + x\,,\ \ldots\,,$$

we will write

$$2x\,,\ 3x\,,\ 4x\,,\ \ldots\,,$$

respectively. For example

$$4\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 0 \end{pmatrix}.$$

Defining $4x = x + x + x + x$ is fine for integer multiples, but does not help us make sense of $\frac{1}{3}x$. For the different types of vectors above, you can probably

guess how to multiply a vector by a scalar. For example

$$\frac{1}{3}\begin{pmatrix}1\\1\\0\end{pmatrix}=\begin{pmatrix}\frac{1}{3}\\\frac{1}{3}\\0\end{pmatrix}.$$

A very special vector can be produced from any vector of any kind by scalar multiplying any vector by the number 0. This is called the *zero vector* and is usually denoted simply 0. This gives five very different kinds of zero from the 5 different kinds of vectors in examples A-E above.

(A) $0(3)=0$ (The zero number)

(B) $0\begin{pmatrix}1\\1\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}$ (The zero 3-vector)

(C) $0\,(1+x-2x^2+3x^3)=0$ (The zero polynomial)

(D) $0\big(1+x-\frac{1}{2!}x^2+\frac{1}{3!}x^3+\cdots\big)=0+0x+0x^2+0x^3+\cdots$ (The zero power series)

(E) $0\,(e^x)=0$ (The zero function)

In any given situation that you plan to describe using vectors, you need to decide on a way to add and scalar multiply vectors. In summary:

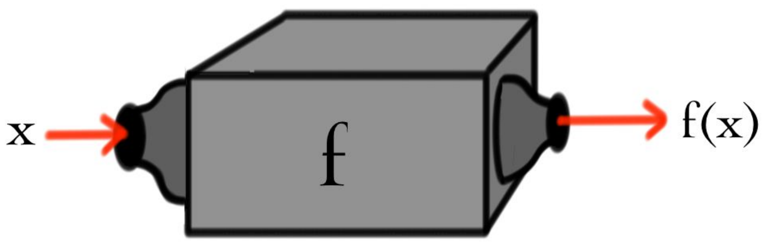
Vectors are things you can add and scalar multiply.

Examples of kinds of vectors:

- numbers
- n-vectors
- 2nd order polynomials
- polynomials
- power series
- functions with a certain domain

1.3 What are Linear Functions?

In calculus classes, the main subject of investigation was the rates of change of functions. In linear algebra, functions will again be the focus of your attention, but functions of a very special type. In precalculus you were perhaps encouraged to think of a function as a machine f into which one may feed a real number. For each input x this machine outputs a single real number $f(x)$.



In linear algebra, the functions we study will have vectors (of some type) as both inputs and outputs. We just saw that vectors are objects that can be added or scalar multiplied—a very general notion—so the functions we are going to study will look novel at first. So things don't get too abstract, here are five questions that can be rephrased in terms of functions of vectors.

Example 3 (Questions involving Functions of Vectors in Disguise)

(A) What number x satisfies $10x = 3$?

(B) What 3-vector u satisfies[4] $\begin{pmatrix}1\\1\\0\end{pmatrix} \times u = \begin{pmatrix}0\\1\\1\end{pmatrix}$?

(C) What polynomial p satisfies $\int_{-1}^{1} p(y)dy = 0$ and $\int_{-1}^{1} yp(y)dy = 1$?

(D) What power series $f(x)$ satisfies $x\frac{d}{dx}f(x) - 2f(x) = 0$?

[4] The cross product appears in this equation.

(E) What number x satisfies $4x^2 = 1$?

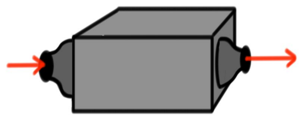
All of these are of the form

($\star$) What vector X satisfies $f(X) = B$?

with a function[5] f known, a vector B known, and a vector X unknown.

The machine needed for part (A) is as in the picture below.

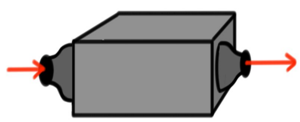
x



$10x$

This is just like a function f from calculus that takes in a number x and spits out the number $10x$. (You might write $f(x) = 10x$ to indicate this). For part (B), we need something more sophisticated.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

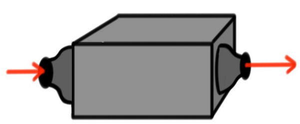


$$\begin{pmatrix} z \\ -z \\ y-x \end{pmatrix},$$

The inputs and outputs are both 3-vectors. The output is the cross product of the input with... how about you complete this sentence to make sure you understand.

The machine needed for example (C) looks like it has just one input and two outputs; we input a polynomial and get a 2-vector as output.

p



$$\begin{pmatrix} \int_{-1}^{1} p(y)dy \\ \int_{-1}^{1} yp(y)dy \end{pmatrix}.$$

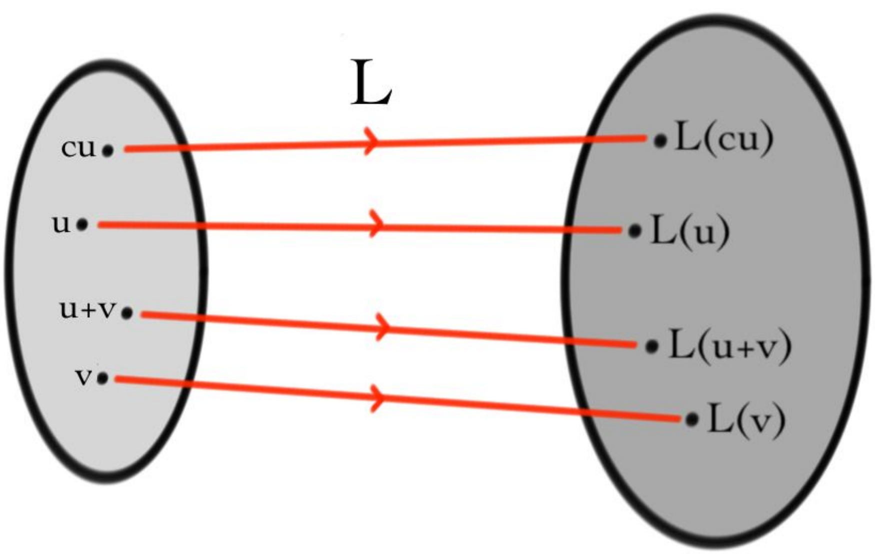
This example is important because it displays an important feature; *the inputs for this function are functions.*

[5]In math terminology, each question is asking for the level set of f corresponding to B.

While this sounds complicated, linear algebra is the study of simple functions of vectors; its time to describe the essential characteristics of linear functions.

Let's use the letter L to denote an arbitrary linear function and think again about vector addition and scalar multiplication. Also, suppose that v and u are vectors and c is a number. Since L is a function from vectors to vectors, if we input u into L, the output $L(u)$ will also be some sort of vector. The same goes for $L(v)$. (And remember, our input and output vectors might be something other than stacks of numbers!) Because vectors are things that can be added and scalar multiplied, $u+v$ and cu are also vectors, and so they can be used as inputs. The essential characteristic of linear functions is what can be said about $L(u+v)$ and $L(cu)$ in terms of $L(u)$ and $L(v)$.

Before we tell you this essential characteristic, ruminate on this picture.



The "blob" on the left represents all the vectors that you are allowed to input into the function L, the blob on the right denotes the possible outputs, and the lines tell you which inputs are turned into which outputs.[6] A full pictorial description of the functions would require all inputs and outputs

[6]The domain, codomain, and rule of correspondence of the function are represented by the left blog, right blob, and arrows, respectively.

and lines to be explicitly drawn, but we are being diagrammatic; we only drew four of each.

Now think about adding $L(u)$ and $L(v)$ to get yet another vector $L(u) + L(v)$ or of multiplying $L(u)$ by c to obtain the vector $cL(u)$, and placing both on the right blob of the picture above. But wait! Are you certain that these are possible outputs!?

Here's the answer

The key to the whole class, from which everything else follows:

1. Additivity:

$$L(u+v) = L(u) + L(v)\,.$$

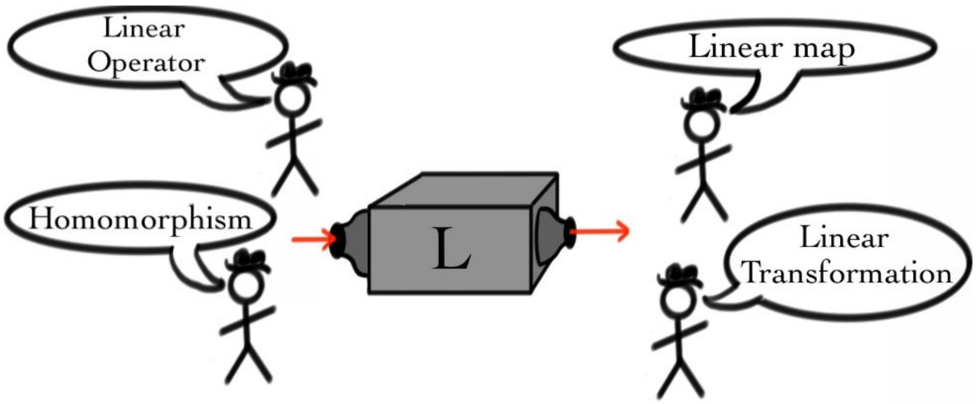
2. Homogeneity:

$$L(cu) = cL(u)\,.$$

Most functions of vectors do not obey this requirement.[7] At its heart, linear algebra is the study of functions that do.

Notice that the additivity requirement says that the function L respects vector addition: *it does not matter if you first add u and v and then input their sum into L, or first input u and v into L separately and then add the outputs.* The same holds for scalar multiplication–try writing out the scalar multiplication version of the italicized sentence. When a function of vectors obeys the additivity and homogeneity properties we say that it is *linear* (this is the "linear" of linear algebra). Together, additivity and homogeneity are called *linearity*. Are there other, equivalent, names for linear functions? yes.

[7] *E.g.:* If $f(x) = x^2$ then $f(1+1) = 4 \neq f(1) + f(1) = 2$. Try any other function you can think of!



$$\boxed{\text{Function} = \text{Transformation} = \text{Operator}}$$

And now for a hint at the power of linear algebra. The questions in examples (A-D) can all be restated as

$$\boxed{Lv = w}$$

where v is an unknown, w a known vector, and L is a known linear transformation. To check that this is true, one needs to know the rules for adding vectors (both inputs and outputs) and then check linearity of L. Solving the equation $Lv = w$ often amounts to solving systems of linear equations, the skill you will learn in Chapter 2.

A great example is the derivative operator.

Example 4 (The derivative operator is linear)
For any two functions $f(x)$, $g(x)$ and any number c, in calculus you probably learnt that the derivative operator satisfies

1. $\frac{d}{dx}(cf) = c\frac{d}{dx}f$,
2. $\frac{d}{dx}(f+g) = \frac{d}{dx}f + \frac{d}{dx}g$.

If we view functions as vectors with addition given by addition of functions and with scalar multiplication given by multiplication of functions by constants, then these familiar properties of derivatives are just the linearity property of linear maps.

Before introducing matrices, notice that for linear maps L we will often write simply Lu instead of $L(u)$. This is because the linearity property of a

linear transformation L means that $L(u)$ can be thought of as multiplying the vector u by the linear operator L. For example, the linearity of L implies that if u, v are vectors and c, d are numbers, then

$$L(cu + dv) = cLu + dLv\,,$$

which feels a lot like the regular rules of algebra for numbers. Notice though, that "uL" makes no sense here.

Remark A sum of multiples of vectors $cu + dv$ is called a *linear combination* of u and v.

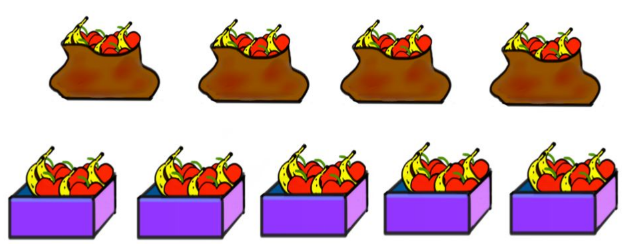
1.4 So, What is a Matrix?

Matrices are linear functions of a certain kind. They appear almost ubiquitously in linear algebra because– and this is the central lesson of introductory linear algebra courses–

> Matrices are the result of organizing information related to linear functions.

This idea will take some time to develop, but we provided an elementary example in Section 1.1. A good starting place to learn about matrices is by studying *systems of linear equations*.

Example 5 A room contains x bags and y boxes of fruit.



Each bag contains 2 apples and 4 bananas and each box contains 6 apples and 8 bananas. There are 20 apples and 28 bananas in the room. Find x and y.

The values are the numbers x and y that simultaneously make both of the following equations true:

$$\begin{array}{rcl} 2\,x + 6\,y &=& 20 \\ 4\,x + 8\,y &=& 28\,. \end{array}$$

Here we have an example of a *System of Linear Equations*.[8] It's a collection of equations in which variables are multiplied by constants and summed, and no variables are multiplied together: There are no powers of variables (like x^2 or y^5), non-integer or negative powers of variables (like $y^{1/7}$ or x^{-3}), and no places where variables are multiplied together (like xy).

Reading homework: problem 1

Information about the fruity contents of the room can be stored two ways:

(i) In terms of the number of apples and bananas.

(ii) In terms of the number of bags and boxes.

Intuitively, knowing the information in one form allows you to figure out the information in the other form. Going from (ii) to (i) is easy: If you knew there were 3 bags and 2 boxes it would be easy to calculate the number of apples and bananas, and doing so would have the feel of multiplication (containers times fruit per container). In the example above we are required to go the other direction, from (i) to (ii). This feels like the opposite of multiplication, *i.e.*, division. Matrix notation will make clear what we are "multiplying" and "dividing" by.

The goal of Chapter 2 is to efficiently solve systems of linear equations. Partly, this is just a matter of finding a better notation, but one that hints at a deeper underlying mathematical structure. For that, we need rules for adding and scalar multiplying 2-vectors;

$$c\begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} cx \\ cy \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x' \\ y' \end{pmatrix} := \begin{pmatrix} x + x' \\ y + y' \end{pmatrix}\,.$$

[8] Perhaps you can see that both lines are of the form $Lu = v$ with $u = \begin{pmatrix} x \\ y \end{pmatrix}$ an unknown, $v = 20$ in the first line, $v = 28$ in the second line, and L different functions in each line? We give the typical less sophisticated description in the text above.

Writing our fruity equations as an equality between 2-vectors and then using these rules we have:

$$\left.\begin{array}{r}2\,x+6\,y=20\\4\,x+8\,y=28\end{array}\right\}\iff\begin{pmatrix}2x+6y\\4x+8y\end{pmatrix}=\begin{pmatrix}20\\28\end{pmatrix}\iff x\begin{pmatrix}2\\4\end{pmatrix}+y\begin{pmatrix}6\\8\end{pmatrix}=\begin{pmatrix}20\\28\end{pmatrix}.$$

Now we introduce a function which takes in 2-vectors[9] and gives out 2-vectors. We denote it by an array of numbers called a *matrix* .

$$\textbf{The function } \begin{pmatrix}2&6\\4&8\end{pmatrix} \textbf{ is defined by } \begin{pmatrix}2&6\\4&8\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}:=x\begin{pmatrix}2\\4\end{pmatrix}+y\begin{pmatrix}6\\8\end{pmatrix}.$$

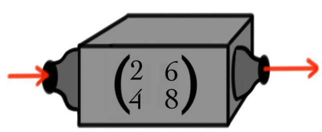
A similar definition applies to matrices with different numbers and sizes.

Example 6 (A bigger matrix)

$$\begin{pmatrix}1&0&3&4\\5&0&3&4\\-1&6&2&5\end{pmatrix}\begin{pmatrix}x\\y\\z\\w\end{pmatrix}:=x\begin{pmatrix}1\\5\\-1\end{pmatrix}+y\begin{pmatrix}0\\0\\6\end{pmatrix}+z\begin{pmatrix}3\\3\\2\end{pmatrix}+w\begin{pmatrix}4\\4\\5\end{pmatrix}.$$

Viewed as a machine that inputs and outputs 2-vectors, our 2×2 matrix does the following:

$$\begin{pmatrix}x\\y\end{pmatrix}$$



$$\begin{pmatrix}2x+6y\\4x+8y\end{pmatrix}.$$

Our fruity problem is now rather concise.

Example 7 (This time in purely mathematical language):

What vector $\begin{pmatrix}x\\y\end{pmatrix}$ satisfies $\begin{pmatrix}2&6\\4&8\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}20\\28\end{pmatrix}$?

[9]To be clear, we will use the term 2-vector to refer to stacks of two numbers such as $\begin{pmatrix}7\\11\end{pmatrix}$. If we wanted to refer to the vectors x^2+1 and x^3-1 (recall that polynomials are vectors) we would say "consider the two vectors x^3-1 and x^2+1". We apologize through giggles for the possibility of the phrase "two 2-vectors."

This is of the same $Lv = w$ form as our opening examples. The matrix encodes fruit per container. The equation is roughly fruit per container times number of containers equals fruit. To solve for number of containers we want to somehow "divide" by the matrix.

Another way to think about the above example is to remember the rule for multiplying a matrix times a vector. If you have forgotten this, you can actually guess a good rule by making sure the matrix equation is the same as the system of linear equations. This would require that

$$\begin{pmatrix} 2 & 6 \\ 4 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} 2x + 6y \\ 4x + 8y \end{pmatrix}$$

Indeed this is an example of the general rule that you have probably seen before

$$\begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} px + qy \\ rx + sy \end{pmatrix} = x \begin{pmatrix} p \\ r \end{pmatrix} + y \begin{pmatrix} q \\ s \end{pmatrix} .$$

Notice, that the second way of writing the output on the right hand side of this equation is very useful because it tells us what all possible outputs a matrix times a vector look like – they are just sums of the columns of the matrix multiplied by scalars. The set of all possible outputs of a matrix times a vector is called the **column space** (it is also the image of the linear function defined by the matrix).

Reading homework: problem 2

Multiplication by a matrix is an example of a *Linear Function*, because it takes one vector and turns it into another in a "linear" way. Of course, we can have much larger matrices if our system has more variables.

Matrices in Space!

Thus matrices can be viewed as linear functions. The statement of this for the matrix in our fruity example is as follows.

1. $\begin{pmatrix} 2 & 6 \\ 4 & 8 \end{pmatrix} \lambda \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} 2 & 6 \\ 4 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ and

2. $\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\left[\begin{pmatrix}x\\ y\end{pmatrix}+\begin{pmatrix}x'\\ y'\end{pmatrix}\right]=\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}+\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\begin{pmatrix}x'\\ y'\end{pmatrix}.$

These equalities can be verified using the rules we introduced so far.

Example 8 Verify that $\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}$ is a linear operator.

The matrix-function is homogeneous if the expressions on the left hand side and right hand side of the first equation are indeed equal.

$$\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\left[\lambda\begin{pmatrix}a\\ b\end{pmatrix}\right]=\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\begin{pmatrix}\lambda a\\ \lambda b\end{pmatrix}=\lambda a\begin{pmatrix}2\\ 4\end{pmatrix}+\lambda b\begin{pmatrix}6\\ 8\end{pmatrix}$$

$$=\begin{pmatrix}2\lambda a\\ 4\lambda a\end{pmatrix}+\begin{pmatrix}6bc\\ 8bc\end{pmatrix}=\underline{\begin{pmatrix}2\lambda a+6\lambda b\\ 4\lambda a+8\lambda b\end{pmatrix}}$$

while

$$\lambda\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix}=c\left[a\begin{pmatrix}2\\ 4\end{pmatrix}+b\begin{pmatrix}6\\ 8\end{pmatrix}\right]=\lambda\left[\begin{pmatrix}2a\\ 4a\end{pmatrix}+\begin{pmatrix}6b\\ 8b\end{pmatrix}\right]$$

$$=\lambda\begin{pmatrix}2a+6b\\ 4a+8b\end{pmatrix}=\underline{\begin{pmatrix}2\lambda a+6\lambda b\\ 4\lambda a+8\lambda b\end{pmatrix}}.$$

The underlined expressions are identical, so the matrix is homogeneous.

The matrix-function is additive if the left and right side of the second equation are indeed equal.

$$\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\left[\begin{pmatrix}a\\ b\end{pmatrix}+\begin{pmatrix}c\\ d\end{pmatrix}\right]=\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\begin{pmatrix}a+c\\ b+d\end{pmatrix}=(a+c)\begin{pmatrix}2\\ 4\end{pmatrix}+(b+d)\begin{pmatrix}6\\ 8\end{pmatrix}$$

$$=\begin{pmatrix}2(a+c)\\ 4(a+c)\end{pmatrix}+\begin{pmatrix}6(b+d)\\ 8(b+d)\end{pmatrix}=\underline{\begin{pmatrix}2a+2c+6b+6d\\ 4a+4c+8b+8d\end{pmatrix}}$$

which we need to compare to

$$\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix}+\begin{pmatrix}2 & 6\\ 4 & 8\end{pmatrix}\begin{pmatrix}c\\ d\end{pmatrix}=a\begin{pmatrix}2\\ 4\end{pmatrix}+b\begin{pmatrix}6\\ 8\end{pmatrix}+c\begin{pmatrix}2\\ 4\end{pmatrix}+d\begin{pmatrix}6\\ 8\end{pmatrix}$$

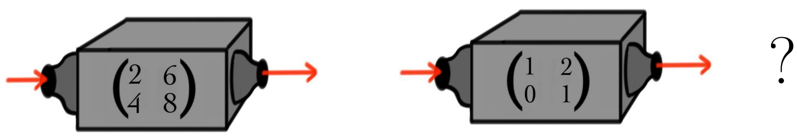
$$=\begin{pmatrix}2a\\ 4a\end{pmatrix}+\begin{pmatrix}6b\\ 8b\end{pmatrix}+\begin{pmatrix}2c\\ 4c\end{pmatrix}+\begin{pmatrix}6d\\ 8d\end{pmatrix}=\underline{\begin{pmatrix}2a+2c+6b+6d\\ 4a+4c+8b+8d\end{pmatrix}}.$$

Thus multiplication by a matrix is additive and homogeneous, and so it is, by definition, linear.

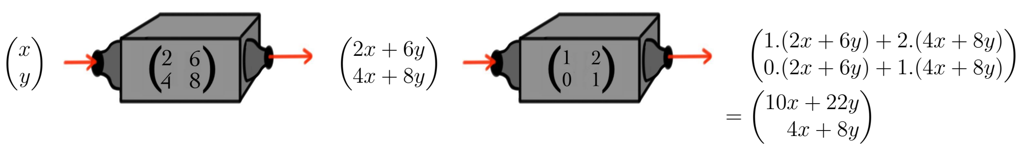
We have come full circle; matrices are just examples of the kinds of linear operators that appear in algebra problems like those in section 1.3. Any equation of the form $Mv = w$ with M a matrix, and v, w n-vectors is called a *matrix equation*. Chapter 2 is about efficiently solving systems of linear equations, or equivalently matrix equations.

1.4.1 Matrix Multiplication is Composition of Functions

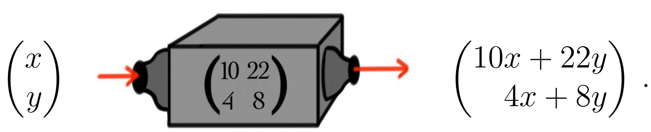
What would happen if we placed two of our expensive machines end to end?



The output of the first machine would be fed into the second.



Notice that the same final result could be achieved with a single machine:



There is a simple matrix notation for this called *matrix multiplication*

$$\begin{pmatrix}1&2\\0&1\end{pmatrix}\begin{pmatrix}2&6\\4&8\end{pmatrix}=\begin{pmatrix}10&22\\4&8\end{pmatrix}.$$

Try review problem 6 to learn more about matrix multiplication.

In the language[10] of functions, if

$$f : U \longrightarrow V \quad \text{and} \quad g : V \longrightarrow W$$

[10]The notation $h : A \to B$ means that h is a function with domain A and codomain B. See the webwork background set3 if you are unfamiliar with this notation or these terms.

the new function obtained by plugging the outputs if f into g is called $g \circ f$,

$$g \circ f : U \longrightarrow W$$

where

$$(g \circ f)(u) = g(f(u)) .$$

This is called the *composition of functions*. Matrix multiplication is the tool required for computing the composition of linear functions.

1.4.2 The Matrix Detour

Linear algebra is about linear functions, not matrices. The following presentation is meant to get you thinking about this idea constantly throughout the course.

> Matrices only get involved in linear algebra when certain notational choices are made.

To exemplify, lets look at the derivative operator again.

Example 9 of how matrices come into linear algebra.
Consider the equation

$$\left(\frac{d}{dx} + 2\right) f = x + 1$$

where f is unknown (the place where solutions should go) and the linear differential operator $\frac{d}{dx} + 2$ is understood to take in quadratic functions (of the form $ax^2 + bx + c$) and give out other quadratic functions.

Let's simplify the way we denote the quadratic functions; we will

$$\text{denote } \; ax^2 + bx + c \; \text{ as } \; \begin{pmatrix} a \\ b \\ c \end{pmatrix}_B .$$

The subscript B serves to remind us of our particular notational convention; we will compare to another notational convention later. With the convention B we can say

$$\left(\frac{d}{dx} + 2\right) \begin{pmatrix} a \\ b \\ c \end{pmatrix}_B = \left(\frac{d}{dx} + 2\right) (ax^2 + bx + c)$$

$$= (2ax + b) + (2ax^2 + 2bx + 2c) = 2ax^2 + (2a + 2b)x + (b + 2c)$$

$$= \begin{pmatrix} 2a \\ 2a + 2b \\ b + 2c \end{pmatrix}_B = \left[\begin{pmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right]_B .$$

That is, our notational convention for quadratic functions has induced a notation for the differential operator $\frac{d}{dx} + 2$ as a matrix. We can use this notation to change the way that the following two equations say exactly the same thing.

$$\left(\frac{d}{dx} + 2 \right) f = x + 1 \Leftrightarrow \left[\begin{pmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \right]_B = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}_B .$$

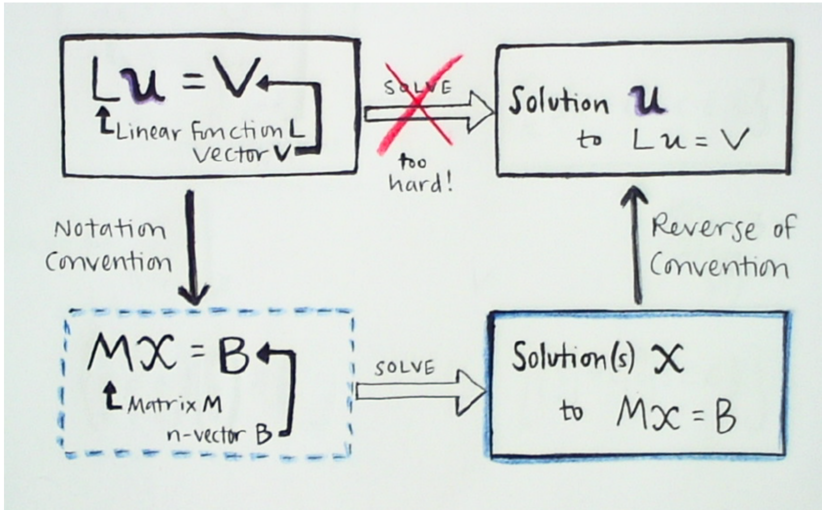
Our notational convention has served as an organizing principle to yield the system of equations

$$\begin{aligned} 2a &= 0 \\ 2a + 2b &= 1 \\ b + 2c &= 1 \end{aligned}$$

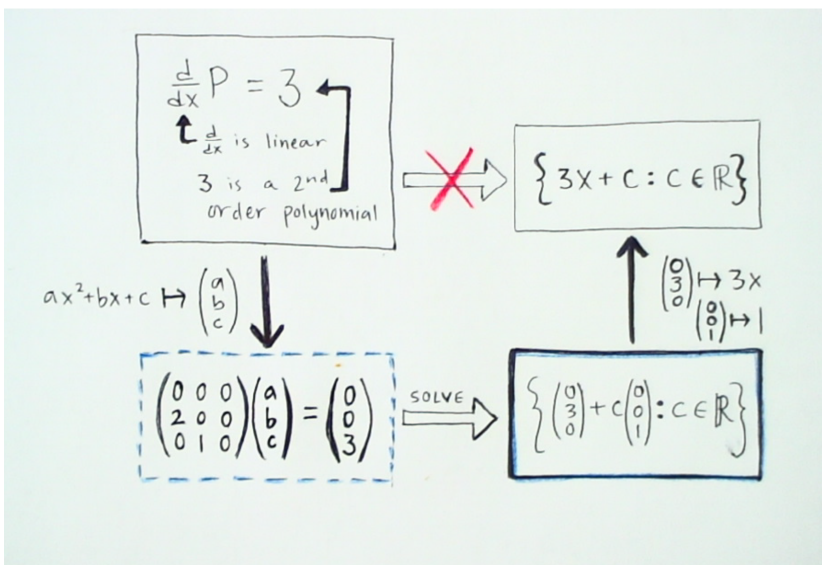
with solution $\begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{4} \end{pmatrix}_B$, where the subscript B is used to remind us that this stack of numbers encodes the vector $\frac{1}{2}x + \frac{1}{4}$, which is indeed the solution to our equation since, substituting for f yields the true statement $\left(\frac{d}{dx} + 2\right)\left(\frac{1}{2}x + \frac{1}{4}\right) = x + 1$.

It would be nice to have a systematic way to rewrite any linear equation as an equivalent matrix equation. It will be a little while before we can learn to organize information in a way generalizable to all linear equations, but keep this example in mind throughout the course.

The general idea is presented in the picture below; sometimes a linear equation is too hard to solve as is, but by organizing information and reformulating the equation as a matrix equation the process of finding solutions becomes tractable.



A simple example with the knowns (L and V are $\frac{d}{dx}$ and 3, respectively) is shown below, although the detour is unnecessary in this case since you know how to anti-differentiate.



To drive home the point that we are not studying matrices but rather linear functions, and that those linear functions can be represented as matrices under certain notational conventions, consider how changeable the notational conventions are.

Example 10 of how a different matrix comes into the same linear algebra problem.

Another possible notational convention is to

$$\text{denote } a+bx+cx^2 \text{ as } \begin{pmatrix} a \\ b \\ c \end{pmatrix}_{B'} .$$

With this alternative notation

$$\left(\frac{d}{dx}+2\right)\begin{pmatrix} a \\ b \\ c \end{pmatrix}_{B'} = \left(\frac{d}{dx}+2\right)(a+bx+cx^2)$$

$$= (b+2cx)+(2a+2bx+2cx^2) = (2a+b)+(2b+2c)x+2cx^2$$

$$= \begin{pmatrix} 2a+b \\ 2b+2c \\ 2c \end{pmatrix}_{B'} = \left[\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right]_{B'} .$$

Notice that we have obtained *a different matrix for the same linear function*. The equation we started with

$$\left(\frac{d}{dx}+2\right)f = x+1 \Leftrightarrow \left[\begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix}\right]_{B'} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}_{B'}$$

$$\begin{aligned} 2a+b &= 1 \\ \Leftrightarrow \quad 2b+2c &= 1 \\ 2c &= 0 \end{aligned}$$

has the solution $\begin{pmatrix} \frac{1}{4} \\ \frac{1}{2} \\ 0 \end{pmatrix}$. Notice that we have obtained *a different 3-vector for the same vector*, since in the notational convention B' this 3-vector represents $\frac{1}{4}+\frac{1}{2}x$.

One linear function can be represented (denoted) by a huge variety of matrices. The representation only depends on how vectors are denoted as n-vectors.

1.5 Review Problems

You probably have already noticed that understanding sets, functions and basic logical operations is a must to do well in linear algebra. Brush up on these skills by trying these background webwork problems:

Logic	1
Sets	2
Functions	3
Equivalence Relations	4
Proofs	5

Each chapter also has reading and skills WeBWorK problems:

Webwork:	Reading problems	1, 2

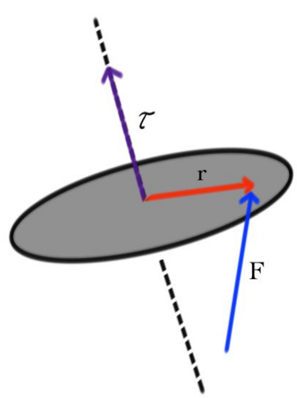
Probably you will spend most of your time on the following review questions:

1. Problems A, B, and C of example 3 can all be written as $Lv = w$ where

$$L : V \longrightarrow W\,,$$

(read this as L maps the set of vectors V to the set of vectors W). For each case write down the sets V and W where the vectors v and w come from.

2. Torque is a measure of "rotational force". It is a vector whose direction is the (preferred) axis of rotation. Upon applying a force F on an object at point r the torque τ is the cross product $r \times F = \tau$:



Remember that the *cross product* of two 3-vectors is given by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \times \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} := \begin{pmatrix} yz' - zy' \\ zx' - xz' \\ xy' - yx' \end{pmatrix}.$$

Indeed, 3-vectors are special, usually vectors an only be added, not multiplied.

Lets find the force F (a vector) one must apply to a wrench lying along the vector $r = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ ft, to produce a torque $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ ft lb:

(a) Find a solution by writing out this equation with $F = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

(Hint: Guess and check that a solution with $a = 0$ exists).

(b) Add $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ to your solution and check that the result is a solution.

(c) Give a physics explanation of why there can be two solutions, and argue that there are, in fact, infinitely many solutions.

(d) Set up a system of three linear equations with the three components of F as the variables which describes this situation. What happens if you try to solve these equations by substitution?

3. The function $P(t)$ gives gas prices (in units of dollars per gallon) as a function of t the year (in A.D. or C.E.), and $g(t)$ is the gas consumption rate measured in gallons per year by a driver as a function of their age. The function g is certainly different for different people. Assuming a lifetime is 100 years, what function gives the total amount spent on gas during the lifetime of an individual born in an arbitrary year t? Is the operator that maps g to this function linear?

4. The differential equation (DE)

$$\frac{d}{dt} f = 2f$$

says that the rate of change of f is proportional to f. It describes exponential growth because the exponential function

$$f(t) = f(0)e^{2t}$$

satisfies the DE for any number $f(0)$. The number 2 in the DE is called the constant of proportionality. A similar DE

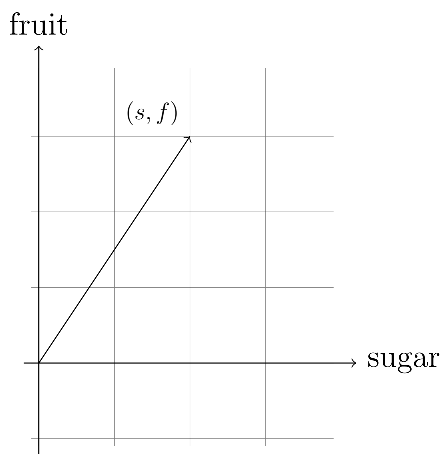
$$\frac{d}{dt}f = \frac{2}{t}f$$

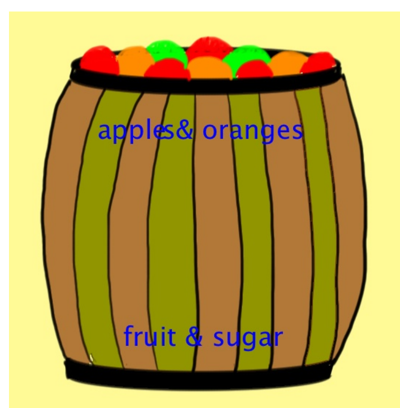
has a time-dependent "constant of proportionality".

(a) Do you think that the second DE describes exponential growth?

(b) Write both DEs in the form $Df = 0$ with D a linear operator.

5. Pablo is a nutritionist who knows that oranges always have twice as much sugar as apples. When considering the sugar intake of schoolchildren eating a barrel of fruit, he represents the barrel like so:





Find a linear operator relating Pablo's representation to the "everyday" representation in terms of the number of apples and number of oranges. Write your answer as a matrix.

Hint: Let λ represent the amount of sugar in each apple.



Hint



6. *Matrix Multiplication:* Let M and N be matrices

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } N = \begin{pmatrix} e & f \\ g & h \end{pmatrix},$$

and v the vector

$$v = \begin{pmatrix} x \\ y \end{pmatrix}.$$

If we first apply N and then M to v we obtain the vector MNv.

(a) Show that the composition of matrices MN is also a linear operator.

(b) Write out the components of the matrix product MN in terms of the components of M and the components of N. *Hint*: use the general rule for multiplying a 2-vector by a 2×2 matrix.

(c) Try to answer the following common question, "Is there any sense in which these rules for matrix multiplication are unavoidable, or are they just a notation that could be replaced by some other notation?"

(d) Generalize your multiplication rule to 3×3 matrices.

7. *Diagonal matrices:* A matrix M can be thought of as an array of numbers m^i_j, known as matrix entries, or matrix components, where i and j index row and column numbers, respectively. Let

$$M = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \left(m^i_j\right).$$

Compute m^1_1, m^1_2, m^2_1 and m^2_2.

The matrix entries m^i_i whose row and column numbers are the same are called the *diagonal* of M. Matrix entries m^i_j with $i \neq j$ are called *off-diagonal*. How many diagonal entries does an $n \times n$ matrix have? How many off-diagonal entries does an $n \times n$ matrix have?

If all the off-diagonal entries of a matrix vanish, we say that the matrix is diagonal. Let

$$D = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \quad \text{and} \quad D' = \begin{pmatrix} \lambda' & 0 \\ 0 & \mu' \end{pmatrix}.$$

Are these matrices diagonal and why? Use the rule you found in problem 6 to compute the matrix products DD' and $D'D$. What do you observe? Do you think the same property holds for arbitrary matrices? What about products where only one of the matrices is diagonal?

(p.s. Diagonal matrices play a special role in in the study of matrices in linear algebra. Keep an eye out for this special role.)

8. Find the linear operator that takes in vectors from n-space and gives out vectors from n-space in such a way that

   (a) whatever you put in, you get exactly the same thing out as what you put in. Show that it is unique. Can you write this operator as a matrix?

   (b) whatever you put in, you get exactly the same thing out as when you put something else in. Show that it is unique. Can you write this operator as a matrix?

   Hint: To show something is unique, it is usually best to begin by pretending that it isn't, and then showing that this leads to a nonsensical conclusion. In mathspeak–*proof by contradiction*.

9. Consider the set $S = \{*, \star, \#\}$. It contains just 3 elements, and has no ordering; $\{*, \star, \#\} = \{\#, \star, *\}$ *etc*. (In fact the same is true for $\{1, 2, 3\} = \{2, 3, 1\}$ *etc*, although we could make this an *ordered set* using $3 > 2 > 1$.)

   (i) Invent a function with domain $\{*, \star, \#\}$ and codomain $\mathbb{R}$. (Remember that the *domain* of a function is the set of all its allowed inputs and the *codomain* (or *target space*) is the set where the outputs can live. A function is specified by assigning exactly one codomain element to each element of the domain.)

   (ii) Choose an ordering on $\{*, \star, \#\}$, and then use it to write your function from part (i) as a triple of numbers.

   (iii) Choose a new ordering on $\{*, \star, \#\}$ and then write your function from part (i) as a triple of numbers.

(iv) Your answers for parts (ii) and (iii) are different yet represent the same function – explain!

2 Systems of Linear Equations

2.1 Gaussian Elimination

Systems of linear equations can be written as matrix equations. Now you will learn an efficient algorithm for (maximally) simplifying a system of linear equations (or a matrix equation) – Gaussian elimination.

2.1.1 Augmented Matrix Notation

Efficiency demands a new notation, called an *augmented matrix*, which we introduce via examples:

The linear system

$$\left\{\begin{array}{rcrcr} x & + & y & = & 27 \\ 2x & - & y & = & 0\,, \end{array}\right.$$

is denoted by the augmented matrix

$$\left(\begin{array}{rr|r} 1 & 1 & 27 \\ 2 & -1 & 0 \end{array}\right).$$

This notation is simpler than the matrix one,

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 27 \\ 0 \end{pmatrix},$$

although all three of the above denote the same thing.



Augmented Matrix Notation



Another interesting rewriting is

$$x\begin{pmatrix}1\\2\end{pmatrix}+y\begin{pmatrix}1\\-1\end{pmatrix}=\begin{pmatrix}27\\0\end{pmatrix}.$$

This tells us that we are trying to find the combination of the vectors $\begin{pmatrix}1\\2\end{pmatrix}$ and $\begin{pmatrix}1\\-1\end{pmatrix}$ adds up to $\begin{pmatrix}27\\0\end{pmatrix}$; the answer is "clearly" $9\begin{pmatrix}1\\2\end{pmatrix}+18\begin{pmatrix}1\\-1\end{pmatrix}$.

Here is a larger example. The system

$$\begin{array}{rcl}1x+3y+2z+0w&=&9\\6x+2y+0z-2w&=&0\\-1x+0y+1z+1w&=&3\,,\end{array}$$

is denoted by the augmented matrix

$$\left(\begin{array}{rrrr|r}1&3&2&0&9\\6&2&0&-2&0\\-1&0&1&1&3\end{array}\right),$$

which is equivalent to the matrix equation

$$\begin{pmatrix}1&3&2&0\\6&2&0&-2\\-1&0&1&1\end{pmatrix}\begin{pmatrix}x\\y\\z\\w\end{pmatrix}=\begin{pmatrix}9\\0\\3\end{pmatrix}.$$

Again, we are trying to find which combination of the columns of the matrix adds up to the vector on the right hand side.

For the the general case of r linear equations in k unknowns, the number of equations is the number of rows r in the augmented matrix, and the number of columns k in the matrix left of the vertical line is the number of unknowns, giving an augmented matrix of the form

$$\left(\begin{array}{cccc|c}a^1_1&a^1_2&\cdots&a^1_k&b^1\\a^2_1&a^2_2&\cdots&a^2_k&b^2\\\vdots&\vdots&&\vdots&\vdots\\a^r_1&a^r_2&\cdots&a^r_k&b^r\end{array}\right).$$

Entries left of the divide carry two indices; subscripts denote column number and superscripts row number. We emphasize, the superscripts here do *not* denote exponents. Make sure you can write out the system of equations and the associated matrix equation for any augmented matrix.

Reading homework: problem 1

We now have three ways of writing the same question. Let's put them side by side as we solve the system by strategically adding and subtracting equations. We will not tell you the motivation for this particular series of steps yet, but let you develop some intuition first.

Example 11 (How matrix equations and augmented matrices change in elimination)

$$\left.\begin{array}{rcrcr} x & + & y & = & 27 \\ 2x & - & y & = & 0 \end{array}\right\} \Leftrightarrow \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 27 \\ 0 \end{pmatrix} \Leftrightarrow \left(\begin{array}{rr|r} 1 & 1 & 27 \\ 2 & -1 & 0 \end{array}\right).$$

With the first equation replaced by the sum of the two equations this becomes

$$\left.\begin{array}{rcrcr} 3x & + & 0 & = & 27 \\ 2x & - & y & = & 0 \end{array}\right\} \Leftrightarrow \begin{pmatrix} 3 & 0 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 27 \\ 0 \end{pmatrix} \Leftrightarrow \left(\begin{array}{rr|r} 3 & 0 & 27 \\ 2 & -1 & 0 \end{array}\right).$$

Let the new first equation be the old first equation divided by 3:

$$\left.\begin{array}{rcrcr} x & + & 0 & = & 9 \\ 2x & - & y & = & 0 \end{array}\right\} \Leftrightarrow \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 9 \\ 0 \end{pmatrix} \Leftrightarrow \left(\begin{array}{rr|r} 1 & 0 & 9 \\ 2 & -1 & 0 \end{array}\right).$$

Replace the second equation by the second equation minus two times the first equation:

$$\left.\begin{array}{rcrcr} x & + & 0 & = & 9 \\ 0 & - & y & = & -18 \end{array}\right\} \Leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 9 \\ -18 \end{pmatrix} \Leftrightarrow \left(\begin{array}{rr|r} 1 & 0 & 9 \\ 0 & -1 & -18 \end{array}\right).$$

Let the new second equation be the old second equation divided by -1:

$$\left.\begin{array}{rcrcr} x & + & 0 & = & 9 \\ 0 & + & y & = & 18 \end{array}\right\} \Leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 9 \\ 18 \end{pmatrix} \Leftrightarrow \left(\begin{array}{rr|r} 1 & 0 & 9 \\ 0 & 1 & 18 \end{array}\right).$$

Did you see what the strategy was? To *eliminate* y from the first equation and then *eliminate* x from the second. The result was the solution to the system.

Here is the big idea: Everywhere in the instructions above we can replace the word "equation" with the word "row" and interpret them as telling us what to do with the augmented matrix instead of the system of equations. Performed systemically, the result is the **Gaussian elimination** algorithm.

2.1.2 Equivalence and the Act of Solving

We now introduce the symbol $\sim$ which is called "tilde" but should be read as "is (row) equivalent to" because at each step the augmented matrix changes by an operation on its rows but its solutions do not. For example, we found above that

$$\left(\begin{array}{cc|c} 1 & 1 & 27 \\ 2 & -1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 9 \\ 2 & -1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 9 \\ 0 & 1 & 18 \end{array}\right).$$

The last of these augmented matrices is our favorite!



Equivalence Example



Setting up a string of equivalences like this is a means of solving a system of linear equations. This is the main idea of section 2.1.3. This next example hints at the main trick:

Example 12 (Using Gaussian elimination to solve a system of linear equations)

$$\left.\begin{array}{rcr} x+\ y & = & 5 \\ x+2y & = & 8 \end{array}\right\} \Leftrightarrow \left(\begin{array}{cc|c} 1 & 1 & 5 \\ 1 & 2 & 8 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & 5 \\ 0 & 1 & 3 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \end{array}\right) \Leftrightarrow \left\{\begin{array}{rcr} x+0 & = & 2 \\ 0+y & = & 3 \end{array}\right.$$

Note that in going from the first to second augmented matrix, we used the top left 1 to make the bottom left entry zero. For this reason we call the top left entry a pivot. Similarly, to get from the second to third augmented matrix, the bottom right entry (before the divide) was used to make the top right one vanish; so the bottom right entry is also called a pivot.

This name *pivot* is used to indicate the matrix entry used to "zero out" the other entries in its column; the pivot is the number used to eliminate another number in its column.

2.1.3 Reduced Row Echelon Form

For a system of two linear equations, the goal of Gaussian elimination is to convert the part of the augmented matrix left of the dividing line into the matrix

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

called the *Identity Matrix*, since this would give the simple statement of a solution $x = a, y = b$. The same goes for larger systems of equations for which the identity matrix I has 1's along its diagonal and all off-diagonal entries vanish:

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Reading homework: problem 2

For many systems, it is not possible to reach the identity in the augmented matrix via Gaussian elimination. In any case, a certain version of the matrix that has the maximum number of components eliminated is said to be the Row Reduced Echelon Form (RREF).

Example 13 (Redundant equations)

$$\left.\begin{array}{rcrcr} x & + & y & = & 2 \\ 2x & + & 2y & = & 4 \end{array}\right\} \Leftrightarrow \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 2 & 4 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right) \Leftrightarrow \left\{\begin{array}{rcrcr} x & + & y & = & 2 \\ 0 & + & 0 & = & 0 \end{array}\right.$$

This example demonstrates if one equation is a multiple of the other the identity matrix can not be a reached. This is because the first step in elimination will make the second row a row of zeros. Notice that solutions still exists $(1, 1)$ is a solution. The last augmented matrix here is in RREF; no more than two components can be eliminated.

Example 14 (Inconsistent equations)

$$\left.\begin{array}{rcrcr} x & + & y & = & 2 \\ 2x & + & 2y & = & 5 \end{array}\right\} \Leftrightarrow \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 2 & 5 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 0 & 1 \end{array}\right) \Leftrightarrow \left\{\begin{array}{rcrcr} x & + & y & = & 2 \\ 0 & + & 0 & = & 1 \end{array}\right.$$

This system of equation has a solution if there exists two numbers x, and y such that $0 + 0 = 1$. That is a tricky way of saying there are no solutions. The last form of the augmented matrix here is the RREF.

Example 15 (Silly order of equations)
A robot might make this mistake:

$$\left.\begin{array}{rcrcr} 0x & + & y & = & -2 \\ x & + & y & = & 7 \end{array}\right\} \Leftrightarrow \left(\begin{array}{cc|r} 0 & 1 & -2 \\ 1 & 1 & 7 \end{array}\right) \sim \cdots ,$$

and then give up because the the upper left slot can not function as a pivot since the 0 that lives there can not be used to eliminate the zero below it. Of course, the right thing to do is to change the order of the equations before starting

$$\left.\begin{array}{rcrcr} x & + & y & = & 7 \\ 0x & + & y & = & -2 \end{array}\right\} \Leftrightarrow \left(\begin{array}{cc|r} 1 & 1 & 7 \\ 0 & 1 & -2 \end{array}\right) \sim \left(\begin{array}{cc|r} 1 & 0 & 9 \\ 0 & 1 & -2 \end{array}\right) \Leftrightarrow \left\{\begin{array}{rcrcr} x & + & 0 & = & 9 \\ 0 & + & y & = & -2\,. \end{array}\right.$$

The third augmented matrix above is the RREF of the first and second. That is to say, you can swap rows on your way to RREF.

For larger systems of equations redundancy and inconsistency are the obstructions to obtaining the identity matrix, and hence to a simple statement of a solution in the form $x = a, y = b, \ldots$. What can we do to maximally simplify a system of equations in general? We need to perform operations that simplify our system *without changing its solutions*. Because, exchanging the order of equations, multiplying one equation by a *non-zero* constant or adding equations does not change the system's solutions, we are lead to three operations:

- (Row Swap) Exchange any two rows.
- (Scalar Multiplication) Multiply any row by a non-zero constant.
- (Row Addition) Add one row to another row.

These are called *Elementary Row Operations*, or EROs for short, and are studied in detail in section 2.3. Suppose now we have a general augmented matrix for which the first entry in the first row does not vanish. Then, using just the three EROs, we could[1] then perform the following.

[1]This is a "brute force" algorithm; there will often be more efficient ways to get to RREF.

Algorithm For Obtaining RREF:

- Make the leftmost nonzero entry in the top row 1 by multiplication.
- Then use that 1 as a pivot to eliminate everything below it.
- Then go to the next row and make the leftmost nonzero entry 1.
- Use that 1 as a pivot to eliminate everything below *and above it*!
- Go to the next row and make the leftmost nonzero entry 1... *etc*

In the case that the first entry of the first row is zero, we may first interchange the first row with another row whose first entry is non-vanishing and then perform the above algorithm. If the entire first column vanishes, we may still apply the algorithm on the remaining columns.

Here is a video (with special effects!) of a hand performing the algorithm by hand. When it is done, you should try doing what it does.



`Beginner Elimination`

This algorithm and its variations is known as Gaussian elimination. The endpoint of the algorithm is an augmented matrix of the form

$$\left(\begin{array}{cccccccc|c}
1 & * & 0 & * & 0 & \cdots & 0 & * & b^1 \\
0 & 0 & 1 & * & 0 & \cdots & 0 & * & b^2 \\
0 & 0 & 0 & 0 & 1 & \cdots & 0 & * & b^3 \\
\vdots & \vdots & \vdots & & \vdots & & & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & 1 & * & b^k \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & b^{k+1} \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & b^r
\end{array}\right).$$

This is called *Reduced Row Echelon Form* (RREF). The asterisks denote the possibility of arbitrary numbers (*e.g.*, the second 1 in the top line of example 13).

Learning to perform this algorithm by hand is the first step to learning linear algebra; it will be the primary means of computation for this course. You need to learn it well. So start practicing as soon as you can, and practice often.

The following properties define RREF:

1. In every row the left most non-zero entry is 1 (and is called a pivot).
2. The pivot of any given row is always to the right of the pivot of the row above it.
3. The pivot is the only non-zero entry in its column.

Example 16 (Augmented matrix in RREF)

$$\left(\begin{array}{ccc|c} 1 & 0 & 7 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Example 17 (Augmented matrix NOT in RREF)

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

Actually, this NON-example breaks all three of the rules!

The reason we need the asterisks in the general form of RREF is that not every column need have a pivot, as demonstrated in examples 13 and 16. Here is an example where multiple columns have no pivot:

Example 18 (Consecutive columns with no pivot in RREF)

$$\left.\begin{array}{ccccccccc} x & + & y & + & z & + & 0w & = & 2 \\ 2x & + & 2y & + & 2z & + & 2w & = & 4 \end{array}\right\} \Leftrightarrow \left(\begin{array}{cccc|c} 1 & 1 & 1 & 0 & 2 \\ 2 & 2 & 2 & 1 & 4 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right)$$

$$\Leftrightarrow \left\{\begin{array}{ccccccc} x & + & y & + & z & = & 2 \\ & & & & w & = & 0\,. \end{array}\right.$$

Note that there was no hope of reaching the identity matrix, because of the shape of the augmented matrix we started with.

With some practice, elimination can go quickly. Here is an expert showing you some tricks. If you can't follow him now then come back when you have some more experience and watch again. You are going to need to get really good at this!



Advanced Elimination

It is important that you are able to convert RREF back into a system of equations. The first thing you might notice is that if any of the numbers $b^{k+1}, \ldots, b^r$ in 2.1.3 are non-zero then the system of equations is inconsistent and has no solutions. Our next task is to extract all possible solutions from an RREF augmented matrix.

2.1.4 Solution Sets and RREF

RREF is a maximally simplified version of the original system of equations in the following sense:

- As many coefficients of the variables as possible are 0.
- As many coefficients of the variables as possible are 1.

It is easier to read off solutions from the maximally simplified equations than from the original equations, even when there are infinitely many solutions.

Example 19 (Standard approach from a system of equations to the solution set)

$$\left.\begin{array}{rcrcrcrcr} x & + & y & & & + & 5w & = & 1 \\ & & y & & & + & 2w & = & 6 \\ & & & & z & + & 4w & = & 8 \end{array}\right\} \Leftrightarrow \left(\begin{array}{cccc|c} 1 & 1 & 0 & 5 & 1 \\ 0 & 1 & 0 & 2 & 6 \\ 0 & 0 & 1 & 4 & 8 \end{array}\right) \sim \left(\begin{array}{cccc|r} 1 & 0 & 0 & 3 & -5 \\ 0 & 1 & 0 & 2 & 6 \\ 0 & 0 & 1 & 4 & 8 \end{array}\right)$$

$$\Leftrightarrow \left\{\begin{array}{rcrcrcrcr} x & & & & & + & 3w & = & -5 \\ & & y & & & + & 2w & = & 6 \\ & & & & z & + & 4w & = & 8 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{rcrcr} x & = & -5 & - & 3w \\ y & = & 6 & - & 2w \\ z & = & 8 & - & 4w \\ w & = & & & w \end{array}\right\}$$

$$\Leftrightarrow \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} -5 \\ 6 \\ 8 \\ 0 \end{pmatrix} + w \begin{pmatrix} -3 \\ -2 \\ -4 \\ 1 \end{pmatrix}.$$

There is one solution for each value of w, so the solution set is

$$\left\{ \begin{pmatrix} -5 \\ 6 \\ 8 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -3 \\ -2 \\ -4 \\ 1 \end{pmatrix} : \alpha \in \mathbb{R} \right\}.$$

Here is a verbal description of the preceeding example of the *standard approach*. We say that x, y, and z are *pivot variables* because they appeared with a pivot coefficient in RREF. Since w never appears with a pivot coefficient, it is not a pivot variable. In the second line we put all the pivot variables on one side and all the *non-pivot variables* on the other side and added the trivial equation $w = w$ to obtain a system that allowed us to easily read off solutions.

The Standard Approach To Solution Sets

1. Write the augmented matrix.

2. Perform EROs to reach RREF.

3. Express the pivot variables in terms of the non-pivot variables.

There are always exactly enough non-pivot variables to index your solutions. In any approach, the variables which are not expressed in terms of the other variables are called *free variables*. The standard approach is to use the non-pivot variables as free variables.

Non-standard approach: solve for w in terms of z and substitute into the other equations. You now have an expression for each component in terms of z. But why pick z instead of y or x? (or $x + y$?) The standard approach not only feels natural, but is *canonical*, meaning that everyone will get the same RREF and hence choose the same variables to be free. However, it is important to remember that so long as their *set* of solutions is the same, any two choices of free variables is fine. (You might think of this as the difference between using Google Maps™ or Mapquest™; although their maps may look different, the place ⟨home *sic*⟩ they are describing is the same!)

When you see an RREF augmented matrix with two columns that have no pivot, you know there will be two free variables.

Example 20 (Standard approach, multiple free variables)

$$\begin{pmatrix} 1 & 0 & 7 & 0 & 4 \\ 0 & 1 & 3 & 4 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \Leftrightarrow \left\{ \begin{array}{rl} x \quad + 7z & = 4 \\ y + 3z + 4w & = 1 \end{array} \right\}$$

$$\Leftrightarrow \left\{ \begin{array}{rcrcrcr} x & = & 4 & - & 7z & & \\ y & = & 1 & - & 3z & - & 4w \\ z & = & & & z & & \\ w & = & & & & & w \end{array} \right\} \Leftrightarrow \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 0 \\ 0 \end{pmatrix} + z \begin{pmatrix} -7 \\ -3 \\ 1 \\ 0 \end{pmatrix} + w \begin{pmatrix} 0 \\ -4 \\ 0 \\ 1 \end{pmatrix}$$

so the solution set is

$$\left\{ \begin{pmatrix} 4 \\ 1 \\ 0 \\ 0 \end{pmatrix} + z \begin{pmatrix} -7 \\ -3 \\ 1 \\ 0 \end{pmatrix} + w \begin{pmatrix} 0 \\ -4 \\ 0 \\ 1 \end{pmatrix} : z, w \in \mathbb{R} \right\}.$$



From RREF to a Solution Set



You can imagine having three, four, or fifty-six non-pivot columns and the same number of free variables indexing your solutions set. In general a solution set to a system of equations with n free variables will be of the form

$$\{x^P + \mu_1 x_1^H + \mu_2 x_2^H + \cdots + \mu_n x_n^H : \mu_1, \ldots, \mu_n \in \mathbb{R}\}.$$

The parts of these solutions play special roles in the associated matrix equation. This will come up again and again long after we complete this discussion of basic calculation methods, so we will use the general language of linear algebra to give names to these parts now.

Definition: A **homogeneous solution** to a linear equation $Lx = v$, with L and v known is a vector x^H such that $Lx^H = 0$ where 0 is the zero vector.

> If you have a particular solution x^P to a linear equation and add a sum of multiples of homogeneous solutions to it you obtain another particular solution.

Particular and Homogeneous Solutions

Check now that the parts of the solutions with free variables as coefficients from the previous examples are homogeneous solutions, and that by adding a homogeneous solution to a particular solution one obtains a solution to the matrix equation. This will come up over and over again. As an example without matrices, consider the differential equation $\frac{d^2}{dx^2}f = 3$. A particular solution is $\frac{3}{2}x^2$ while x and 1 are homogeneous solutions. The solution set is $\{\frac{3}{2}x^2 + ax + c1 \ : \ a, b \in \mathbb{R}\}$. You can imagine similar differential equations with more homogeneous solutions.

You need to become very adept at reading off solutions sets of linear systems from the RREF of their augmented matrix; it is a basic skill for linear algebra, and we will continue using it up to the last page of the book!

Worked examples of Gaussian elimination

2.2 Review Problems

Webwork:

Reading problems	1, 2
Augmented matrix	6
2×2 systems	7, 8, 9, 10, 11, 12
3×2 systems	13, 14
3×3 systems	15, 16, 17

1. State whether the following augmented matrices are in RREF and compute their solution sets.

$$\left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 3 & 1 \\ 0 & 1 & 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 & 2 & 0 \end{array}\right),$$

$$\left(\begin{array}{cccccc|c} 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right),$$

$$\left(\begin{array}{ccccccc|c} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 & 0 & 2 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right).$$

2. Solve the following linear system:

$$\begin{array}{rcl} 2x_1 + 5x_2 - 8x_3 + 2x_4 + 2x_5 &=& 0 \\ 6x_1 + 2x_2 - 10x_3 + 6x_4 + 8x_5 &=& 6 \\ 3x_1 + 6x_2 + 2x_3 + 3x_4 + 5x_5 &=& 6 \\ 3x_1 + 1x_2 - 5x_3 + 3x_4 + 4x_5 &=& 3 \\ 6x_1 + 7x_2 - 3x_3 + 6x_4 + 9x_5 &=& 9 \end{array}$$

   Be sure to set your work out carefully with equivalence signs $\sim$ between each step, labeled by the row operations you performed.

3. Check that the following two matrices are row-equivalent:

$$\left(\begin{array}{ccc|c} 1 & 4 & 7 & 10 \\ 2 & 9 & 6 & 0 \end{array}\right) \text{ and } \left(\begin{array}{ccc|c} 0 & -1 & 8 & 20 \\ 4 & 18 & 12 & 0 \end{array}\right).$$

   Now remove the third column from each matrix, and show that the resulting two matrices (shown below) are row-equivalent:

$$\left(\begin{array}{cc|c} 1 & 4 & 10 \\ 2 & 9 & 0 \end{array}\right) \text{ and } \left(\begin{array}{cc|c} 0 & -1 & 20 \\ 4 & 18 & 0 \end{array}\right).$$

   Now remove the fourth column from each of the original two matrices, and show that the resulting two matrices, viewed as augmented matrices (shown below) are row-equivalent:

$$\left(\begin{array}{cc|c} 1 & 4 & 7 \\ 2 & 9 & 6 \end{array}\right) \text{ and } \left(\begin{array}{cc|c} 0 & -1 & 8 \\ 4 & 18 & 12 \end{array}\right).$$

   Explain why row-equivalence is never affected by removing columns.

4. Check that the system of equations corresponding to the augmented matrix

$$\left(\begin{array}{cc|c} 1 & 4 & 10 \\ 3 & 13 & 9 \\ 4 & 17 & 20 \end{array}\right)$$

has no solutions. If you remove one of the rows of this matrix, does the new matrix have any solutions? In general, can row equivalence be affected by removing rows? Explain why or why not.

5. Explain why the linear system has no solutions:

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 1 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 6 \end{array}\right)$$

For which values of k does the system below have a solution?

$$\begin{array}{rcrcrcr} x & - & 3y & & & = & 6 \\ x & & & + & 3z & = & -3 \\ 2x & + & ky & + & (3-k)z & = & 1 \end{array}$$



Hint



6. Show that the RREF of a matrix is unique. (Hint: Consider what happens if the same augmented matrix had two different RREFs. Try to see what happens if you removed columns from these two RREF augmented matrices.)

7. Another method for solving linear systems is to use row operations to bring the augmented matrix to Row Echelon Form (REF as opposed to RREF). In REF, the pivots are not necessarily set to one, and we only require that all entries left of the pivots are zero, not necessarily entries above a pivot. Provide a counterexample to show that row echelon form is not unique.

   Once a system is in row echelon form, it can be solved by "back substitution." Write the following row echelon matrix as a system of equations, then solve the system using back-substitution.

$$\left(\begin{array}{ccc|c} 2 & 3 & 1 & 6 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 3 & 3 \end{array}\right)$$

8. Show that this pair of augmented matrices are row equivalent, assuming $ad - bc \neq 0$:
$$\left(\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & \frac{de-bf}{ad-bc} \\ 0 & 1 & \frac{af-ce}{ad-bc} \end{array}\right)$$

9. Consider the augmented matrix:
$$\left(\begin{array}{cc|c} 2 & -1 & 3 \\ -6 & 3 & 1 \end{array}\right).$$
Give a *geometric* reason why the associated system of equations has no solution. (Hint, plot the three vectors given by the columns of this augmented matrix in the plane.) Given a general augmented matrix
$$\left(\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right),$$
can you find a condition on the numbers a, b, c and d that corresponds to the geometric condition you found?

10. A relation $\sim$ on a set of objects U is an *equivalence relation* if the following three properties are satisfied:
    - Reflexive: For any $x \in U$, we have $x \sim x$.
    - Symmetric: For any $x, y \in U$, if $x \sim y$ then $y \sim x$.
    - Transitive: For any x, y and $z \in U$, if $x \sim y$ and $y \sim z$ then $x \sim z$.

    Show that row equivalence of matrices is an example of an equivalence relation.

    (For a discussion of equivalence relations, see Homework 0, Problem 4)



Hint



11. Equivalence of augmented matrices does not come from equality of their solution sets. Rather, we define two matrices to be equivalent if one can be obtained from the other by elementary row operations.

    Find a pair of augmented matrices that are not row equivalent but do have the same solution set.

2.3 Elementary Row Operations

Elementary row operations are systems of linear equations relating the old and new rows in Gaussian elimination:

Example 21 (Keeping track of EROs with equations between rows)
We refer to the new kth row as R'_k and the old kth row as R_k.

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 7 \\ 2 & 0 & 0 & 4 \\ 0 & 0 & 1 & 4 \end{array}\right) \overset{\substack{R'_1=0R_1+\ R_2+0R_3 \\ R'_2=\ R_1+0R_2+0R_3 \\ R'_3=0R_1+0R_2+\ R_3}}{\sim} \left(\begin{array}{ccc|c} 2 & 0 & 0 & 4 \\ 0 & 1 & 1 & 7 \\ 0 & 0 & 1 & 4 \end{array}\right) \qquad \begin{pmatrix} R'_1 \\ R'_2 \\ R'_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix}$$

$$\overset{\substack{R'_1=\frac{1}{2}R_1+0R_2+0R_3 \\ R'_2=0R_1+\ R_2+0R_3 \\ R'_3=0R_1+0R_2+\ R_3}}{\sim} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 7 \\ 0 & 0 & 1 & 4 \end{array}\right) \qquad \begin{pmatrix} R'_1 \\ R'_2 \\ R'_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix}$$

$$\overset{\substack{R'_1=\ R_1+0R_2+0R_3 \\ R'_2=0R_1+\ R_2-\ R_3 \\ R'_3=0R_1+0R_2+\ R_3}}{\sim} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{array}\right) \qquad \begin{pmatrix} R'_1 \\ R'_2 \\ R'_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix}$$

On the right, we have listed the relations between old and new rows in matrix notation.

Reading homework: problem 3

2.3.1 EROs and Matrices

Interestingly, the matrix that describes the relationship between old and new rows performs the corresponding ERO on the augmented matrix.

Example 22 (Performing EROs with Matrices)

$$\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}\left(\begin{array}{ccc|c}0&1&1&7\\2&0&0&4\\0&0&1&4\end{array}\right)=\left(\begin{array}{ccc|c}2&0&0&4\\0&1&1&7\\0&0&1&4\end{array}\right)$$

$$\wr$$

$$\begin{pmatrix}\frac{1}{2}&0&0\\0&1&0\\0&0&1\end{pmatrix}\left(\begin{array}{ccc|c}2&0&0&4\\0&1&1&7\\0&0&1&4\end{array}\right)=\left(\begin{array}{ccc|c}1&0&0&2\\0&1&1&7\\0&0&1&4\end{array}\right)$$

$$\wr$$

$$\begin{pmatrix}1&0&0\\0&1&-1\\0&0&1\end{pmatrix}\left(\begin{array}{ccc|c}1&0&0&2\\0&1&1&7\\0&0&1&4\end{array}\right)=\left(\begin{array}{ccc|c}1&0&0&2\\0&1&0&3\\0&0&1&4\end{array}\right)$$

Here we have multiplied the augmented matrix with the matrices that acted on rows listed on the right of example 21.

Realizing EROs as matrices allows us to give a concrete notion of "dividing by a matrix"; we can now perform manipulations on both sides of an equation in a familiar way:

Example 23 (Undoing A in $Ax = b$ slowly, for $A = 6 = 3 \cdot 2$)

$$\begin{array}{rrcr} & 6x & = & 12\\ \Leftrightarrow & 3^{-1}6x & = & 3^{-1}12\\ \Leftrightarrow & 2x & = & 4\\ \Leftrightarrow & 2^{-1}2x & = & 2^{-1}\ 4\\ \Leftrightarrow & 1x & = & 2\end{array}$$

The matrices corresponding to EROs undo a matrix step by step.

Example 24 (Undoing A in $Ax = b$ slowly, for $A = M = ...$)

$$
\begin{array}{rcl}
& \begin{pmatrix}0&1&1\\2&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = & \begin{pmatrix}7\\4\\4\end{pmatrix}\\
\Leftrightarrow & \begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}0&1&1\\2&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = & \begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}7\\4\\4\end{pmatrix}\\
\Leftrightarrow & \begin{pmatrix}2&0&0\\0&1&1\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = & \begin{pmatrix}4\\7\\4\end{pmatrix}\\
\Leftrightarrow & \begin{pmatrix}\frac{1}{2}&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}2&0&0\\0&1&1\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = & \begin{pmatrix}\frac{1}{2}&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}4\\7\\4\end{pmatrix}\\
\Leftrightarrow & \begin{pmatrix}1&0&0\\0&1&1\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = & \begin{pmatrix}2\\7\\4\end{pmatrix}\\
\Leftrightarrow & \begin{pmatrix}1&0&0\\0&1&-1\\0&0&1\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&1\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = & \begin{pmatrix}1&0&0\\0&1&-1\\0&0&1\end{pmatrix}\begin{pmatrix}2\\7\\4\end{pmatrix}\\
\Leftrightarrow & \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = & \begin{pmatrix}2\\3\\4\end{pmatrix}.
\end{array}
$$

This is another way of thinking about Gaussian elimination which feels more like elementary algebra in the sense that you "do something to both sides of an equation" until you have a solution.

2.3.2 Recording EROs in $(M|I)$

Just as we put together $3^{-1}2^{-1} = 6^{-1}$ to get a single thing to apply to both sides of $6x = 12$ to undo 6, we should put together multiple EROs to get a single thing that undoes our matrix. To do this, augment by the identity matrix (not just a single column) and then perform Gaussian elimination. There is no need to write the EROs as systems of equations or as matrices while doing this.

Example 25 (Collecting EROs that undo a matrix)

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right).$$

As we changed the left side from the matrix M to the identity matrix, the right side changed from the identity matrix to the matrix which undoes M.

Example 26 (Checking that one matrix undoes another)

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If the matrices are composed in the opposite order, the result is the same.

$$\begin{pmatrix} 0 & 1 & 1 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Whenever the product of two matrices $MN = I$, we say that N is the inverse of M or $N = M^{-1}$. Conversely M is the inverse of N; $M = N^{-1}$.

In abstract generality, let M be some matrix and, as always, let I stand for the identity matrix. Imagine the process of performing elementary row operations to bring M to the identity matrix:

$$(M|I) \sim (E_1M|E_1) \sim (E_2E_1M|E_2E_1) \sim \cdots \sim (I|\cdots E_2E_1)\,.$$

The ellipses "$\cdots$" stand for additional EROs. The result is a product of matrices that form a matrix which undoes M

$$\cdots E_2E_1M = I\,.$$

This is only true if the RREF of M is the identity matrix.

Definition: A matrix M is **invertible** if its RREF is an identity matrix.

How to find M^{-1}

$$(M|I) \sim (I|M^{-1})$$

Much use is made of the fact that invertible matrices can be undone with EROs. To begin with, since each elementary row operation has an inverse,

$$M = E_1^{-1} E_2^{-1} \cdots ,$$

while the inverse of M is

$$M^{-1} = \cdots E_2 E_1 .$$

This is symbolically verified by

$$M^{-1} M = \cdots E_2 E_1 \, E_1^{-1} E_2^{-1} \cdots = \cdots E_2 \, E_2^{-1} \cdots = \cdots = I .$$

Thus, if M is invertible, then M can be expressed as the product of EROs. (The same is true for its inverse.) This has the feel of the fundamental theorem of arithmetic (integers can be expressed as the product of primes) or the fundamental theorem of algebra (polynomials can be expressed as the product of [complex] first order polynomials); EROs are building blocks of invertible matrices.

2.3.3 The Three Elementary Matrices

We now work toward concrete examples and applications. It is surprisingly easy to translate between EROs and matrices that perform EROs. The matrices corresponding to these kinds are close in form to the identity matrix:

- Row Swap: Identity matrix with two rows swapped.
- Scalar Multiplication: Identity matrix with one diagonal entry not 1.
- Row Sum: The identity matrix with one off-diagonal entry not 0.

Example 27 (Correspondences between EROs and their matrices)

- The row swap matrix that swaps the 2nd and 4th row is the identity matrix with the 2nd and 4th row swapped:

$$\begin{pmatrix}1&0&0&0&0\\0&0&0&1&0\\0&0&1&0&0\\0&1&0&0&0\\0&0&0&0&1\end{pmatrix}.$$

- The scalar multiplication matrix that replaces the 3rd row with 7 times the 3rd row is the identity matrix with 7 in the 3rd row instead of 1:

$$\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&7&0\\0&0&0&1\end{pmatrix}.$$

- The row sum matrix that replaces the 4th row with the 4th row plus 9 times the 2nd row is the identity matrix with a 9 in the 4th row, 2nd column:

$$\begin{pmatrix}1&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&1&0&0&0&0\\0&9&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&1&0\\0&0&0&0&0&0&1\end{pmatrix}.$$

We can write an explicit factorization of a matrix into EROs by keeping track of the EROs used in getting to RREF.

Example 28 (Express M from Example 25 as a product of EROs)
Note that in the previous example one of each of the kinds of EROs is used, in the order just given. Elimination looked like

$$M=\begin{pmatrix}0&1&1\\2&0&0\\0&0&1\end{pmatrix}\stackrel{E_1}{\sim}\begin{pmatrix}2&0&0\\0&1&1\\0&0&1\end{pmatrix}\stackrel{E_2}{\sim}\begin{pmatrix}1&0&0\\0&1&1\\0&0&1\end{pmatrix}\stackrel{E_3}{\sim}\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}=I\,,$$

where the EROs matrices are

$$E_1=\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix},\;E_2=\begin{pmatrix}\frac{1}{2}&0&0\\0&1&0\\0&0&1\end{pmatrix},\;E_3=\begin{pmatrix}1&0&0\\0&1&-1\\0&0&1\end{pmatrix}.$$

The inverse of the ERO matrices (corresponding to the description of the reverse row maniplulations)

$$E_1^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; E_2^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; E_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Multiplying these gives

$$\begin{aligned} E_1^{-1}E_2^{-1}E_3^{-1} &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = M\,. \end{aligned}$$

2.3.4 LU, LDU, and $PLDU$ Factorizations

The process of elimination can be stopped halfway to obtain decompositions frequently used in large computations in sciences and engineering. The first half of the elimination process is to eliminate entries below the diagonal leaving a matrix which is called *upper triangular*. The elementary matrices which perform this part of the elimination are *lower triangular*, as are their inverses. But putting together the upper triangular and lower triangular parts one obtains the so-called LU factorization.

Example 29 (LU factorization)

$$M = \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ -4 & 0 & 9 & 2 \\ 0 & -1 & 1 & -1 \end{pmatrix} \stackrel{E_1}{\sim} \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 3 & 4 \\ 0 & -1 & 1 & -1 \end{pmatrix}$$

$$\stackrel{E_2}{\sim} \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 3 & 1 \end{pmatrix} \stackrel{E_3}{\sim} \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & -3 \end{pmatrix} := U\,,$$

where the EROs and their inverses are

$$E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

$$E_1^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, E_2^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}, E_3^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Applying inverse elementary matrices to both sides of the equality $U = E_3E_2E_1M$ gives $M = E_1^{-1}E_2^{-1}E_3^{-1}U$ or

$$\begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ -4 & 0 & 9 & 2 \\ 0 & -1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & -3 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & -3 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & -3 \end{pmatrix}.$$

This is a lower triangular matrix times an upper triangular matrix.

What if we stop at a different point in elimination? We could multiply rows so that the entries in the diagonal are 1 next. Note that the EROs that do this are diagonal. This gives a slightly different factorization.

Example 30 (LDU factorization building from previous example)

$$M = \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ -4 & 0 & 9 & 2 \\ 0 & -1 & 1 & -1 \end{pmatrix} \stackrel{E_3E_2E_1}{\sim} \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & -3 \end{pmatrix} \stackrel{E_4}{\sim} \begin{pmatrix} 1 & 0 & -\frac{3}{2} & \frac{1}{2} \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & -3 \end{pmatrix}$$

$$\stackrel{E_5}{\sim} \begin{pmatrix} 1 & 0 & -\frac{3}{2} & \frac{1}{2} \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & \frac{4}{3} \\ 0 & 0 & 0 & -3 \end{pmatrix} \stackrel{E_6}{\sim} \begin{pmatrix} 1 & 0 & -\frac{3}{2} & \frac{1}{2} \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & \frac{4}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix} =: U$$

The corresponding elementary matrices are

$$E_4 = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E_5 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E_6 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\frac{1}{3} \end{pmatrix},$$

$$E_4^{-1} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, E_5^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, E_6^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}.$$

The equation $U = E_6E_5E_4E_3E_2E_1M$ can be rearranged as

$$M = (E_1^{-1}E_2^{-1}E_3^{-1})(E_4^{-1}E_5^{-1}E_6^{-1})U.$$

We calculated the product of the first three factors in the previous example; it was named L there, and we will reuse that name here. The product of the next three factors is diagonal and we wil name it D. The last factor we named U (the name means something different in this example than the last example.) The LDU factorization of our matrix is

$$\begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ -4 & 0 & 9 & 2 \\ 0 & -1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\frac{3}{2} & \frac{1}{2} \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & \frac{4}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The LDU factorization of a matrix is a factorization into blocks of EROs of a various types: L is the product of the inverses of EROs which eliminate below the diagonal by row addition, D the product of inverses of EROs which set the diagonal elements to 1 by row multiplication, and U is the product of inverses of EROs which eliminate above the diagonal by row addition.

You may notice that one of the three kinds of row operation is missing from this story. Row exchange may be necessary to obtain RREF. Indeed, so far in this chapter we have been working under the tacit assumption that M can be brought to the identity by just row multiplication and row addition. If row exchange is necessary, the resulting factorization is $LDPU$ where P is the product of inverses of EROs that perform row exchange.

Example 31 ($LDPU$ factorization, building from previous examples)

$$M = \begin{pmatrix} 0 & 1 & 2 & 2 \\ 2 & 0 & -3 & 1 \\ -4 & 0 & 9 & 2 \\ 0 & -1 & 1 & -1 \end{pmatrix} \stackrel{P}{\sim} \begin{pmatrix} 2 & 0 & -3 & 1 \\ 0 & 1 & 2 & 2 \\ -4 & 0 & 9 & 2 \\ 0 & -1 & 1 & -1 \end{pmatrix} \stackrel{E_6E_5E_4E_3E_2E_1}{\sim} L$$

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = P^{-1}$$

$$M = P(E_1^{-1}E_2^{-1}E_3^{-1})(E_4^{-1}E_5^{-1}E_6^{-1})(E_7^{-1})U = PLDU$$

$$\begin{pmatrix} 0 & 1 & 2 & 2 \\ 2 & 0 & -3 & 1 \\ -4 & 0 & 9 & 2 \\ 0 & -1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & -3 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\frac{3}{2} & \frac{1}{2} \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & \frac{4}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2.4 Review Problems

Webwork:

Reading problems	3
Matrix notation	18
LU	19

1. While performing Gaussian elimination on these augmented matrices write the full system of equations describing the new rows in terms of the old rows above each equivalence symbol as in Example 21.

$$\begin{pmatrix} 2 & 2 & 10 \\ 1 & 2 & 8 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 0 & 5 \\ 1 & 1 & -1 & 11 \\ -1 & 1 & 1 & -5 \end{pmatrix}$$

2. Solve the vector equation by applying ERO matrices to each side of the equation to perform elimination. Show each matrix explicitly as in Example 24.

$$\begin{pmatrix} 3 & 6 & 2 \\ 5 & 9 & 4 \\ 2 & 4 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}$$

3. Solve this vector equation by finding the inverse of the matrix through $(M|I) \sim (I|M^{-1})$ and then applying M^{-1} to both sides of the equation.

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 9 \\ 6 \\ 7 \end{pmatrix}$$

4. Follow the method of Examples 29 and 30 to find the LU and LDU factorization of

$$\begin{pmatrix} 3 & 3 & 6 \\ 3 & 5 & 2 \\ 6 & 2 & 5 \end{pmatrix}.$$

5. Multiple matrix equations with the same matrix can be solved simultaneously.

   (a) Solve both systems by performing elimination on just one augmented matrix.

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

(b) Give an interpretation of the columns of M^{-1} in $(M|I) \sim (I|M^{-1})$ in terms of solutions to certain systems of linear equations.

6. How can you convince your fellow students to never make this mistake?

$$\begin{pmatrix} 1 & 0 & 2 & | & 3 \\ 0 & 1 & 2 & | & 3 \\ 2 & 0 & 1 & | & 4 \end{pmatrix} \stackrel{\substack{R_1'=R_1+R_2 \\ R_2'=R_1-R_2 \\ R_3'=R_1+2R_2}}{\sim} \begin{pmatrix} 1 & 1 & 4 & | & 6 \\ 1 & -1 & 0 & | & 0 \\ 1 & 2 & 6 & | & 9 \end{pmatrix}$$

7. Is LU factorization of a matrix unique? Justify your answer.

∞. If you randomly create a matrix by picking numbers out of the blue, it will probably be difficult to perform elimination or factorization; fractions and large numbers will probably be involved. To invent simple problems it is better to start with a simple answer:

(a) Start with any augmented matrix in RREF. Perform EROs to make most of the components non-zero. Write the result on a separate piece of paper and give it to your friend. Ask that friend to find RREF of the augmented matrix you gave them. Make sure they get the same augmented matrix you started with.

(b) Create an upper triangular matrix U and a lower triangular matrix L with only 1s on the diagonal. Give the result to a friend to factor into LU form.

(c) Do the same with an LDU factorization.

2.5 Solution Sets for Systems of Linear Equations

Algebraic equations problems can have multiple solutions. For example $x(x-1)=0$ has two solutions: 0 and 1. By contrast, equations of the form $Ax=b$ with A a linear operator (with scalars the real numbers) have the following property:

If A is a linear operator and b is known, then $Ax=b$ has either

1. One solution

2. No solutions

3. Infinitely many solutions

2.5.1 The Geometry of Solution Sets: Hyperplanes

Consider the following algebra problems and their solutions.

1. $6x = 12$ has one solution: 2.

2a. $0x = 12$ has no solution.

2b. $0x = 0$ has infinitely many solutions; its solution set is $\mathbb{R}$.

In each case the linear operator is a 1×1 matrix. In the first case, the linear operator is invertible. In the other two cases it is not. In the first case, the solution set is a point on the number line, in case 2b the solution set is the whole number line.

Lets examine similar situations with larger matrices: 2×2 matrices.

1. $\begin{pmatrix} 6 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 12 \\ 6 \end{pmatrix}$ has one solution: $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

2a. $\begin{pmatrix} 1 & 3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$ has no solutions.

2bi. $\begin{pmatrix} 1 & 3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$ has solution set $\left\{ \begin{pmatrix} 4 \\ 0 \end{pmatrix} + y \begin{pmatrix} -3 \\ 1 \end{pmatrix} : y \in \mathbb{R} \right\}$.

2bii. $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ has solution set $\left\{ \begin{pmatrix} x \\ y \end{pmatrix} : x, y \in \mathbb{R} \right\}$.

Again, in the first case the linear operator is invertible while in the other cases it is not. When a 2×2 matrix from a matrix equation is not invertible the solution set can be empty, a line in the plane, or the plane itself.

For a system of equations with r equations and k veriables, one can have a number of different outcomes. For example, consider the case of r equations in three variables. Each of these equations is the equation of a plane in three-dimensional space. To find solutions to the system of equations, we look for the common intersection of the planes (if an intersection exists). Here we have five different possibilities:

1. **Unique Solution.** The planes have a unique point of intersection.

2a. **No solutions.** Some of the equations are contradictory, so no solutions exist.

2bi. **Line.** The planes intersect in a common line; any point on that line then gives a solution to the system of equations.

2bii. **Plane.** Perhaps you only had one equation to begin with, or else all of the equations coincide geometrically. In this case, you have a plane of solutions, with two free parameters.

`Planes`

2biii. **All of $\mathbb{R}^3$.** If you start with no information, then any point in $\mathbb{R}^3$ is a solution. There are three free parameters.

In general, for systems of equations with k unknowns, there are $k+2$ possible outcomes, corresponding to the possible numbers (*i.e.*, $0, 1, 2, \ldots, k$) of free parameters in the solutions set, plus the possibility of no solutions. These types of solution sets are hyperplanes, generalizations of planes that behave like planes in $\mathbb{R}^3$ in many ways.

Reading homework: problem 4

`Pictures and Explanation`

2.5.2 Particular Solution + Homogeneous Solutions

Lets look at solution sets again, this time trying to get to their geometric shape. In the standard approach, variables corresponding to columns that do not contain a pivot (after going to reduced row echelon form) are *free*. It is the number of free variables that determines the geometry of the solution set.

Example 32 (Non-pivot variables determine the gemometry of the solution set)

$$\begin{pmatrix}1&0&1&-1\\0&1&-1&1\\0&0&0&0\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}1\\1\\0\end{pmatrix}\Leftrightarrow\left\{\begin{array}{ll}1x_1+0x_2+1x_3-1x_4&=1\\0x_1+1x_2-1x_3+1x_4&=1\\0x_1+0x_2+0x_3+0x_4&=0\end{array}\right.$$

Following the standard approach, express the pivot variables in terms of the non-pivot variables and add "empty equations". Here x_3 and x_4 are non-pivot variables.

$$\left.\begin{array}{rcl}x_1&=&1-x_3+x_4\\x_2&=&1+x_3-x_4\\x_3&=&x_3\\x_4&=&x_4\end{array}\right\}\Leftrightarrow\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}1\\1\\0\\0\end{pmatrix}+x_3\begin{pmatrix}-1\\1\\1\\0\end{pmatrix}+x_4\begin{pmatrix}1\\-1\\0\\1\end{pmatrix}$$

The preferred way to write a solution set S is with set notation;

$$S=\left\{\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}=\begin{pmatrix}1\\1\\0\\0\end{pmatrix}+\mu_1\begin{pmatrix}-1\\1\\1\\0\end{pmatrix}+\mu_2\begin{pmatrix}1\\-1\\0\\1\end{pmatrix}:\mu_1,\mu_2\in\mathbb{R}\right\}.$$

Notice that the first two components of the second two terms come from the non-pivot columns. Another way to write the solution set is

$$S=\left\{x^P+\mu_1x_1^H+\mu_2x_2^H:\mu_1,\mu_2\in\mathbb{R}\right\},$$

where

$$x^P=\begin{pmatrix}1\\1\\0\\0\end{pmatrix},\quad x_1^H=\begin{pmatrix}-1\\1\\1\\0\end{pmatrix},\quad x_2^H=\begin{pmatrix}1\\-1\\0\\1\end{pmatrix}.$$

Here x^P is a *particular solution* while x_1^H and x_2^H are called *homogeneous solutions*. The solution set forms a plane.

2.5.3 Solutions and Linearity

Motivated by example 32, we say that the matrix equation $Mx=v$ has solution set $\{x^P+\mu_1x_1^H+\mu_2x_2^H\,|\,\mu_1,\mu_2\in\mathbb{R}\}$. Recall that matrices are linear operators. Thus

$$M(x^P+\mu_1x_1^H+\mu_2x_2^H)=Mx^P+\mu_1Mx_1^H+\mu_2Mx_2^H=v\,,$$

for *any* $\mu_1, \mu_2 \in \mathbb{R}$. Choosing $\mu_1 = \mu_2 = 0$, we obtain

$$Mx^P = v\,.$$

This is why x^P is an example of a *particular solution*.

Setting $\mu_1 = 1, \mu_2 = 0$, and subtracting $Mx^P = v$ we obtain

$$Mx_1^H = 0\,.$$

Likewise, setting $\mu_1 = 0, \mu_2 = 1$, we obtain

$$Mx_2^H = 0\,.$$

Here x_1^H and x_2^H are examples of what are called *homogeneous* solutions to the system. They *do not* solve the original equation $Mx = v$, but instead its associated *homogeneous equation* $My = 0$.

We have just learnt a fundamental lesson of linear algebra: the solution set to $Ax = b$, where A is a linear operator, consists of a particular solution plus homogeneous solutions.

$$\{\text{Solutions}\} = \{\text{Particular solution} + \text{Homogeneous solutions}\}$$

Example 33 Consider the matrix equation of example 32. It has solution set

$$S = \left\{ \begin{pmatrix}1\\1\\0\\0\end{pmatrix} + \mu_1 \begin{pmatrix}-1\\1\\1\\0\end{pmatrix} + \mu_2 \begin{pmatrix}1\\-1\\0\\1\end{pmatrix} \,:\, \mu_1, \mu_2 \in \mathbb{R} \right\}.$$

Then $Mx^P = v$ says that $\begin{pmatrix}1\\1\\0\\0\end{pmatrix}$ is a solution to the original matrix equation, which is certainly true, but this is not the only solution.

$Mx_1^H = 0$ says that $\begin{pmatrix}-1\\1\\1\\0\end{pmatrix}$ is a solution to the homogeneous equation.

$Mx_2^H = 0$ says that $\begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}$ is a solution to the homogeneous equation.

Notice how adding any multiple of a homogeneous solution to the particular solution yields another particular solution.

Reading homework: problem 4

2.6 Review Problems

Webwork:

Reading problems	4, 5
Solution sets	20, 21, 22
Geometry of solutions	23, 24, 25, 26

1. Write down examples of augmented matrices corresponding to each of the five types of solution sets for systems of equations with three unknowns.

2. Invent a simple linear system that has multiple solutions. Use the standard approach for solving linear systems and a non-standard approach to obtain different descriptions of the solution set. Is the solution set different with different approaches?

3. Let
$$M = \begin{pmatrix} a_1^1 & a_2^1 & \cdots & a_k^1 \\ a_1^2 & a_2^2 & \cdots & a_k^2 \\ \vdots & \vdots & & \vdots \\ a_1^r & a_2^r & \cdots & a_k^r \end{pmatrix} \quad \text{and} \quad x = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^k \end{pmatrix} .$$
Note: x^2 does not denote the square of the column vector x. Instead x^1, x^2, x^3, *etc...*, denote different variables (the components of x); the superscript is an index. Although confusing at first, this notation was invented by Albert Einstein who noticed that quantities like $a_1^2 x^1 + a_2^2 x^2 \cdots + a_k^2 x^k =: \sum_{j=1}^k a_j^2 x^j$, can be written unambiguously as $a_j^2 x^j$. This is called Einstein summation notation. The most important thing to remember is that the index j is a dummy variable,

so that $a^2_j x^j \equiv a^2_i x^i$; this is called "relabeling dummy indices". When dealing with products of sums, you must remember to introduce a new dummy for each term; *i.e.*, $a_i x^i b_i y^i = \sum_i a_i x^i b_i y^i$ does *not* equal $a_i x^i b_j y^j = \big(\sum_i a_i x^i\big)\big(\sum_j b_j y^j\big)$.

Use Einstein summation notation to propose a rule for Mx so that $Mx = 0$ is equivalent to the linear system

$$\begin{array}{ccccc} a^1_1 x^1 & + a^1_2 x^2 & \cdots & + a^1_k x^k & = 0 \\ a^2_1 x^1 & + a^2_2 x^2 & \cdots & + a^2_k x^k & = 0 \\ \vdots & \vdots & & \vdots & \vdots \\ a^r_1 x^1 & + a^r_2 x^2 & \cdots & + a^r_k x^k & = 0 \end{array}$$

Show that your rule for multiplying a matrix by a vector obeys the linearity property.

4. The *standard basis vector* e_i is a column vector with a one in the ith row, and zeroes everywhere else. Using the rule for multiplying a matrix times a vector in problem 3, find a simple rule for multiplying Me_i, where M is the general matrix defined there.

5. If A is a non-linear operator, can the solutions to $Ax = b$ still be written as "general solution=particular solution + homogeneous solutions"? Provide examples.

6. Find a system of equations whose solution set is the walls of a $1 \times 1 \times 1$ cube. (Hint: You may need to restrict the ranges of the variables; could your equations be linear?)

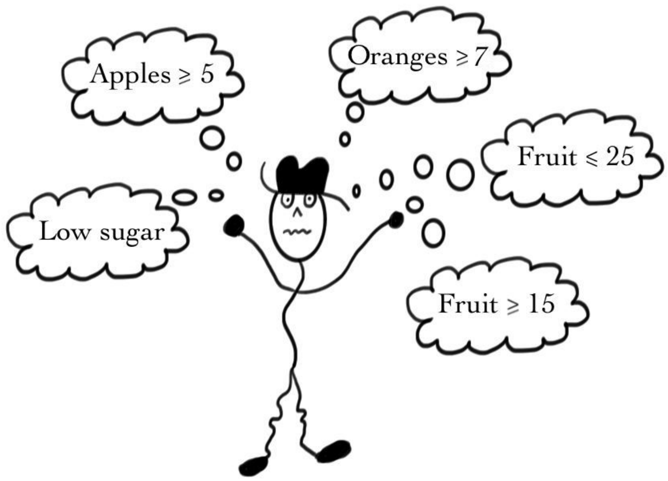
3 The Simplex Method

In Chapter 2, you learned how to handle systems of linear equations. However there are many situations in which inequalities appear instead of equalities. In such cases we are often interested in an optimal solution extremizing a particular quantity of interest. Questions like this are a focus of fields such as mathematical optimization and operations research. For the case where the functions involved are linear, these problems go under the title *linear programming*. Originally these ideas were driven by military applications, but by now are ubiquitous in science and industry. Gigantic computers are dedicated to implementing linear programming methods such as George Dantzig's simplex algorithm–the topic of this chapter.

3.1 Pablo's Problem

Let us begin with an example. Consider again Pablo the nutritionist of problem 5, chapter 1. The Conundrum City school board has employed Pablo to design their school lunch program. Unfortunately for Pablo, their requirements are rather tricky:

Example 34 (Pablo's problem)
The Conundrum City school board is heavily influenced by the local fruit grower's association. They have stipulated that children eat at least 7 oranges and 5 apples per week. Parents and teachers have agreed that eating at least 15 pieces of fruit per week is a good thing, but school janitors argue that too much fruit makes a terrible mess, so that children should eat no more than 25 pieces of fruit per week.



Finally Pablo knows that oranges have twice as much sugar as apples and that apples have 5 grams of sugar each. Too much sugar is unhealthy, so Pablo wants to keep the children's sugar intake as low as possible. How many oranges and apples should Pablo suggest that the school board put on the menu?

This is a rather gnarly word problem. Our first step is to restate it as mathematics, stripping away all the extraneous information:

Example 35 (Pablo's problem restated)
Let x be the number of apples and y be the number of oranges. These must obey

$$x \geq 5 \quad \text{and} \quad y \geq 7\,,$$

to fulfill the school board's politically motivated wishes. The teacher's and parent's fruit requirement means that

$$x + y \geq 15\,,$$

but to keep the canteen tidy

$$x + y \leq 25\,.$$

Now let

$$s = 5x + 10y\,.$$

This linear function of (x, y) represents the grams of sugar in x apples and y oranges. The problem is asking us to minimize s subject to the four linear inequalities listed above.

3.2 Graphical Solutions

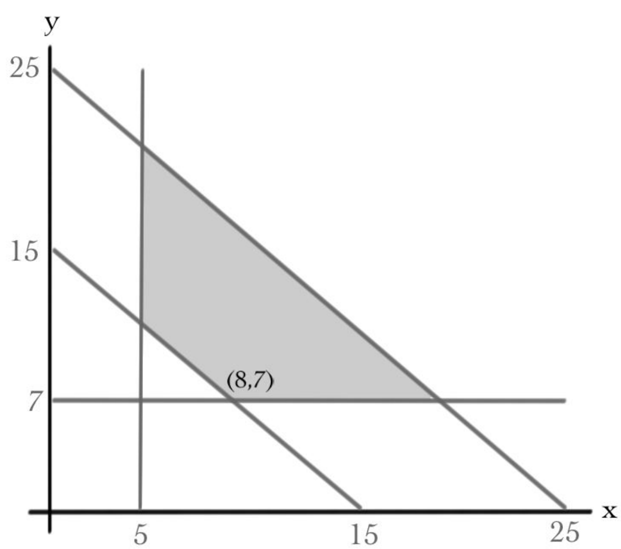
Before giving a more general algorithm for handling this problem and problems like it, we note that when the number of variables is small (preferably 2), a graphical technique can be used.

Inequalities, such as the four given in Pablo's problem, are often called *constraints*, and values of the variables that satisfy these constraints comprise the so-called *feasible region*. Since there are only two variables, this is easy to plot:

Example 36 (Constraints and feasible region) Pablo's constraints are

$$x \geq 5$$
$$y \geq 7$$
$$15 \leq x + y \leq 25\,.$$

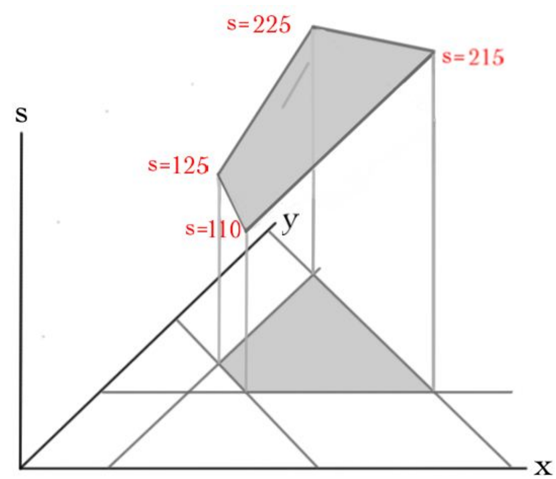
Plotted in the (x, y) plane, this gives:



You might be able to see the solution to Pablo's problem already. Oranges are very sugary, so they should be kept low, thus $y = 7$. Also, the less fruit the better, so the answer had better lie on the line $x + y = 15$. Hence, the answer must be at the vertex $(8, 7)$. Actually this is a general feature

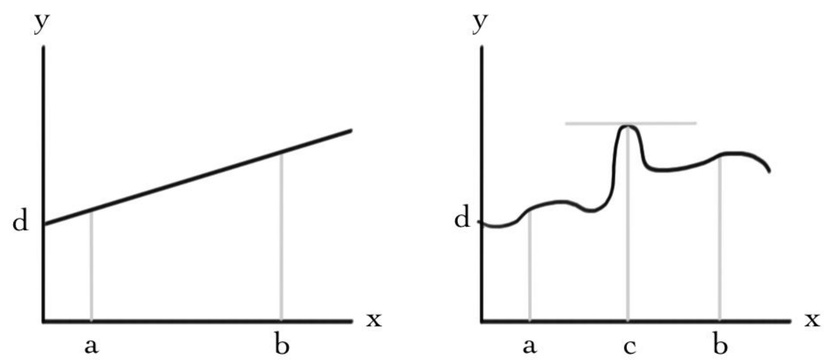
of linear programming problems, the optimal answer must lie at a vertex of the feasible region. Rather than prove this, lets look at a plot of the linear function $s(x, y) = 5x + 10y$.

Example 37 (The sugar function)
Plotting the sugar function requires three dimensions:



The plot of a linear function of two variables is a plane through the origin. Restricting the variables to the feasible region gives some lamina in 3-space. Since the function we want to optimize is linear (and assumedly non-zero), if we pick a point in the middle of this lamina, we can always increase/decrease the function by moving out to an edge and, in turn, along that edge to a corner. Applying this to the above picture, we see that Pablo's best option is 110 grams of sugar a week, in the form of 8 apples and 7 oranges.

It is worthwhile to contrast the optimization problem for a linear function with the non-linear case you may have seen in calculus courses:



Here we have plotted the curve $f(x) = d$ in the case where the function f is linear and non-linear. To optimize f in the interval $[a, b]$, for the linear case we just need to compute and compare the values $f(a)$ and $f(b)$. In contrast, for non-linear functions it is necessary to also compute the derivative df/dx to study whether there are extrema **inside** the interval.

3.3 Dantzig's Algorithm

In simple situations a graphical method might suffice, but in many applications there may be thousands or even millions of variables and constraints. Clearly an algorithm that can be implemented on a computer is needed. The *simplex algorithm* (usually attributed to George Dantzig) provides exactly that. It begins with a standard problem:

Problem 38 Maximize $f(x_1, \ldots, x_n)$ where f is linear, $x_i \geq 0$ $(i = 1, \ldots, n)$ subject to

$$Mx = v\,, \qquad x := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\,,$$

where the $m \times n$ matrix M and $m \times 1$ column vector v are given.

This is solved by arranging the information in an augmented matrix and then applying EROs. To see how this works lets try an example.

Example 39 Maximize $f(x,y,z,w) = 3x - 3y - z + 4w$ subject to constraints

$$\begin{aligned} c_1 &:= \quad x + y + z + w &= 5 \\ c_2 &:= x + 2y + 3z + 2w &= 6\,, \end{aligned}$$

where $x \geq 0$, $y \geq 0$, $z \geq 0$ and $w \geq 0$.

The key observation is this: Suppose we are trying to maximize $f(x_1,\ldots,x_n)$ subject to a constraint $c(x_1,\ldots,x_n) = k$ for some constant k (c and k would be the entries of Mx and v, respectively, in the above). Then we can also try to maximize

$$f(x_1,\ldots,x_n) + \alpha c(x_1,\ldots,x_n)$$

because this is only a constant shift $f \to f + \alpha k$. Choosing α carefully can lead to a simple form for the function we are extremizing.

Example 40 (Setting up an augmented matrix):

Since we are interested in the optimum value of f, we treat it as an additional variable and add one further equation

$$-3x + 3y + z - 4w + f = 0\,.$$

We arrange this equation and the two constraints in an augmented matrix

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 0 & 5 \\ 1 & 2 & 3 & 2 & 0 & 6 \\ \hline -3 & 3 & 1 & -4 & 1 & 0 \end{array}\right) \Leftrightarrow \left\{\begin{array}{rcl} c_1 &=& 5 \\ c_2 &=& 6 \\ f &=& 3x - 3y - z + 4w \end{array}\right. .$$

Keep in mind that the first four columns correspond to the positive variables (x,y,z,w) and that the last row has the information of the function f. The general case is depicted in figure 3.1.

Now the system is written as an augmented matrix where the last row encodes the objective function and the other rows the constraints. Clearly we can perform row operations on the constraint rows since this will not change the solutions to the constraints. Moreover, we can add any amount of the constraint rows to the last row, since this just amounts to adding a constant to the function we want to extremize.

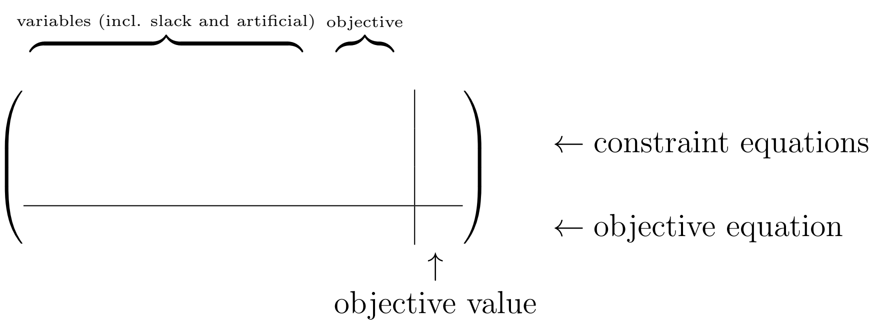


Figure 3.1: Arranging the information of an optimization problem in an augmented matrix.

Example 41 (Performing EROs)

We scan the last row, and notice the (most negative) coefficient -4. Naïvely you might think that this is good because this multiplies the positive variable w and only helps the objective function $f = 4w + \cdots$. However, what this actually means is that the variable w will be positive and thus determined by the constraints. Therefore we want to remove it from the objective function. We can zero out this entry by performing a row operation. For that, either of the first two rows could be used. To decide which, we remember that we still have to solve solve the constraints for variables that are positive. Hence we should try to keep the first two entries in the last column positive. Hence we choose the row which will add the smallest constant to f when we zero out the -4: Look at the last column (where the values of the constraints are stored). We see that adding four times the first row to the last row would zero out the -4 entry but add 20 to f, while adding two times the second row to the last row would also zero out the -4 but only add 12 to f. (You can follow this by watching what happens to the last entry in the last row.) So we perform the latter row operation and obtain the following:

$$\left(\begin{array}{rrrrr|r} 1 & 1 & 1 & 1 & 0 & 5 \\ 1 & 2 & 3 & 2 & 0 & 6 \\ \hline -1 & 7 & 7 & 0 & 1 & 12 \end{array}\right) \quad \left| \quad \begin{array}{rcl} c_1 &=& 5 \\ c_2 &=& 6 \\ f &=& 12 + x - 7y - 7z\,. \end{array}\right.$$

We do not want to undo any of our good work when we perform further row operations, so now we use the second row to zero out all other entries in the fourth column. This is achieved by subtracting half the second row from the first:

$$\left(\begin{array}{rrrrr|r} \frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & 2 \\ 1 & 2 & 3 & 2 & 0 & 6 \\ \hline -1 & 7 & 7 & 0 & 1 & 12 \end{array}\right) \quad \left| \quad \begin{array}{rcl} c_1 - \frac{1}{2}c_2 &=& 2 \\ c_2 &=& 6 \\ f &=& 12 + x - 7y - 7z\,. \end{array}\right.$$

Precisely because we chose the second row to perform our row operations, all entries in the last column remain positive. This allows us to continue the algorithm.

We now repeat the above procedure: There is a -1 in the first column of the last row. We want to zero it out while adding as little to f as possible. This is achieved by adding twice the first row to the last row:

$$\left(\begin{array}{ccccc|c} \frac{1}{2} & 0 & -\frac{1}{2} & 0 & 0 & 2 \\ 1 & 2 & 3 & 2 & 0 & 6 \\ \hline 0 & 7 & 6 & 0 & 1 & 16 \end{array}\right) \quad \left| \quad \begin{array}{rcl} c_1 - \frac{1}{2}c_2 & = & 2 \\ c_2 & = & 6 \\ f & = & 16 - 7y - 6z\,. \end{array}\right.$$

The Dantzig algorithm terminates if all the coefficients in the last row (save perhaps for the last entry which encodes the value of the objective) are positive. To see why we are done, lets write out what our row operations have done in terms of the function f and the constraints (c_1, c_2). First we have

$$f = 16 - 7y - 6z$$

with both y and z positive. Hence to maximize f we should choose $y = 0 = z$. In which case we obtain our optimum value

$$\underline{f = 16\,.}$$

Finally, we check that the constraints can be solved with $y = 0 = z$ and positive (x, w). Indeed, they can by taking $x = 4$, $w = 1$.

3.4 Pablo Meets Dantzig

Oftentimes, it takes a few tricks to bring a given problem into the standard form of example 39. In Pablo's case, this goes as follows.

Example 42 Pablo's variables x and y do not obey $x_i \geq 0$. Therefore define new variables

$$x_1 = x - 5\,, \quad x_2 = y - 7\,.$$

The conditions on the fruit $15 \leq x + y \leq 25$ are inequalities,

$$x_1 + x_2 \geq 3\,, \quad x_1 + x_2 \leq 13\,,$$

so are not of the form $Mx = v$. To achieve this we introduce two new positive variables $x_3 \geq 0$, $x_4 \geq 4$ and write

$$c_1 := x_1 + x_2 - x_3 = 3\,, \quad c_2 := x_1 + x_2 + x_4 = 13\,.$$

These are called *slack variables* because they take up the "slack" required to convert inequality to equality. This pair of equations can now be written as $Mx = v$,

$$\begin{pmatrix} 1 & 1 & -1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ 13 \end{pmatrix} .$$

Finally, Pablo wants to minimize sugar $s = 5x + 10y$, but the standard problem maximizes f. Thus the so-called *objective function* $f = -s + 95 = -5x_1 - 10x_2$. (Notice that it makes no difference whether we maximize $-s$ or $-s + 95$, we choose the latter since it is a linear function of (x_1, x_2).) Now we can build an augmented matrix whose last row reflects the objective function equation $5x_1 + 10x_2 + f = 0$:

$$\left(\begin{array}{ccccc|c} 1 & 1 & -1 & 0 & 0 & 3 \\ 1 & 1 & 0 & 1 & 0 & 13 \\ \hline 5 & 10 & 0 & 0 & 1 & 0 \end{array}\right) .$$

Here it seems that the simplex algorithm already terminates because the last row only has positive coefficients, so that setting $x_1 = 0 = x_2$ would be optimal. However, this does not solve the constraints (for positive values of the slack variables x_3 and x_4). Thus one more (very dirty) trick is needed. We add two more, positive, (so-called) *artificial variables* x_5 and x_6 to the problem which we use to shift each constraint

$$c_1 \to c_1 - x_5 , \qquad c_2 \to c_2 - x_6 .$$

The idea being that for large positive α, the modified objective function

$$f - \alpha x_5 - \alpha x_6$$

is only maximal when the artificial variables vanish so the underlying problem is unchanged. Lets take $\alpha = 10$ (our solution will not depend on this choice) so that our augmented matrix reads

$$\left(\begin{array}{ccccccc|c} 1 & 1 & -1 & 0 & 1 & 0 & 0 & 3 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 13 \\ \hline 5 & 10 & 0 & 0 & 10 & 10 & 1 & 0 \end{array}\right)$$

$$\stackrel{R_3'=R_3-10R_1-10R_2}{\sim} \left(\begin{array}{ccccccc|c} 1 & 1 & -1 & 0 & 1 & 0 & 0 & 3 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 13 \\ \hline -15 & -10 & 10 & -10 & 0 & 0 & 1 & -160 \end{array}\right) .$$

Here we performed one row operation to zero out the coefficients of the artificial variables. Now we are ready to run the simplex algorithm exactly as in section 3.3.

The first row operation uses the 1 in the top of the first column to zero out the most negative entry in the last row:

$$
\left(\begin{array}{ccccccc|c}
1 & 1 & -1 & 0 & 1 & 0 & 0 & 3 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 13 \\ \hline
0 & 5 & -5 & -10 & 15 & 0 & 1 & -115
\end{array}\right)
$$

$$
\stackrel{R_2'=R_2-R_1}{\sim} \left(\begin{array}{ccccccc|c}
1 & 1 & 1 & 0 & 1 & 0 & 0 & 3 \\
0 & 0 & 1 & 1 & -1 & 1 & 0 & 10 \\ \hline
0 & 5 & -5 & -10 & 15 & 0 & 1 & -115
\end{array}\right)
$$

$$
\stackrel{R_3'=R_3+10R_2}{\sim} \left(\begin{array}{ccccccc|c}
1 & 1 & 1 & 0 & 1 & 0 & 0 & 3 \\
0 & 0 & 1 & 1 & -1 & 1 & 0 & 10 \\ \hline
0 & 5 & 5 & 0 & 5 & 10 & 1 & -15
\end{array}\right) .
$$

Now the variables (x_2, x_3, x_5, x_6) have zero coefficients so must be set to zero to maximize f. The optimum value is $f = -15$ so $s = -f + 95 = 110$ exactly as before. Finally, to solve the constraints $x_1 = 3$ and $x_4 = 10$ so that $x = 8$ and $y = 7$ which also agrees with our previous result.

Clearly, performed by hand, the simplex algorithm was slow and complex for Pablo's problem. However, the key point is that it is an algorithm that can be fed to a computer. For problems with many variables, this method is much faster than simply checking all vertices as we did in section 3.2.

3.5 Review Problems

1. Maximize $f(x, y) = 2x + 3y$ subject to the constraints

$$x \geq 0\,, \quad y \geq 0\,, \quad x + 2y \leq 2\,, \quad 2x + y \leq 2\,,$$

   by

   (a) sketching the region in the xy-plane defined by the constraints and then checking the values of f at its corners; and,

   (b) the simplex algorithm (*hint:* introduce slack variables).

2. Conoil operates two wells (well A and well B) in southern Grease (a small Mediterranean country). You have been employed to figure out how many barrels of oil they should pump from each well to maximize

their profit (all of which goes to shareholders, not operating costs). The quality of oil from well A is better than from well B, so is worth 50% more per barrel. The Greasy government cares about the environment and will not allow Conoil to pump in total more than 6 million barrels per year. Well A costs twice as much as well B to operate. Conoil's yearly operating budget is only sufficient to pump at most 10 million barrels from well B per year. Using both a graphical method and then (as a double check) Dantzig's algorithm, determine how many barrels Conoil should pump from each well to maximize their profits.

4

Vectors in Space, n-Vectors

To continue our linear algebra journey, we must discuss n-vectors with an arbitrarily large number of components. The simplest way to think about these is as ordered lists of numbers,

$$a = \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} .$$

Do not be confused by our use of a superscript to label components of a vector. Here a^2 denotes the second component of the vector a, rather than the number a squared!

We emphasize that order matters:

Example 43 (Order of Components Matters)

$$\begin{pmatrix} 7 \\ 4 \\ 2 \\ 5 \end{pmatrix} \neq \begin{pmatrix} 7 \\ 2 \\ 4 \\ 5 \end{pmatrix} .$$

The set of all n-vectors is denoted $\mathbb{R}^n$. As an equation

$$\mathbb{R}^n := \left\{ \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} \middle| \, a^1, \ldots, a^n \in \mathbb{R} \right\} .$$

4.1 Addition and Scalar Multiplication in $\mathbb{R}^n$

A simple but important property of n-vectors is that we can *add* two n-vectors together and *multiply* one n-vector by a scalar:

Definition Given two n-vectors a and b whose components are given by

$$a = \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} \text{ and } b = \begin{pmatrix} b^1 \\ \vdots \\ b^n \end{pmatrix}$$

their *sum* is

$$a + b := \begin{pmatrix} a^1 + b^1 \\ \vdots \\ a^n + b^n \end{pmatrix}.$$

Given a scalar λ, the *scalar multiple*

$$\lambda a := \begin{pmatrix} \lambda a^1 \\ \vdots \\ \lambda a^n \end{pmatrix}.$$

Example 44 Let

$$a = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \text{ and } b = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}.$$

Then, for example,

$$a + b = \begin{pmatrix} 5 \\ 5 \\ 5 \\ 5 \end{pmatrix} \text{ and } 3a - 2b = \begin{pmatrix} -5 \\ 0 \\ 5 \\ 10 \end{pmatrix}.$$

A special vector is the *zero vector*. All of its components are zero:

$$0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} =: 0_n.$$

In Euclidean geometry—the study of $\mathbb{R}^n$ with lengths and angles defined as in section 4.3 —n-vectors are used to label points P and the zero vector labels the origin O. In this sense, the zero vector is the only one with zero magnitude, and the only one which points in no particular direction.

4.2 Hyperplanes

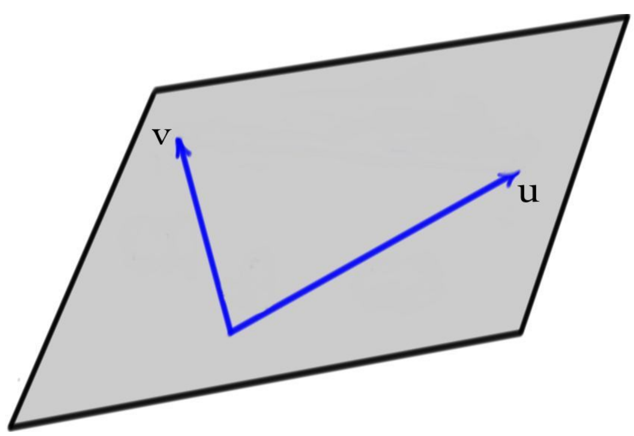
Vectors in $\mathbb{R}^n$ are impossible to visualize unless n is 1,2, or 3. However, familiar objects like lines and planes still make sense for any value of n: The line L along the direction defined by a vector v and through a point P labeled by a vector u can be written as

$$L = \{u + tv \mid t \in \mathbb{R}\} .$$

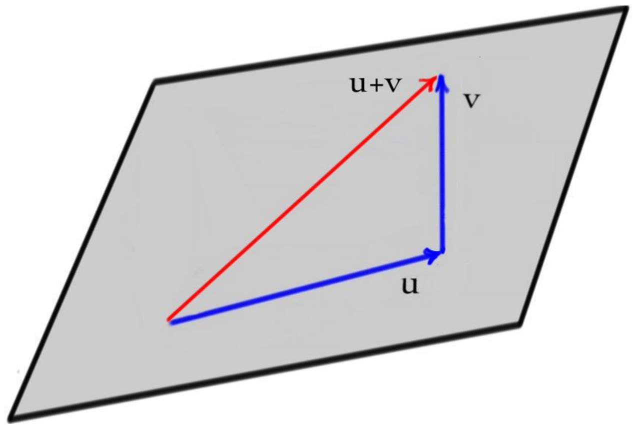
Sometimes, since we know that a point P corresponds to a vector, we will be lazy and just write $L = \{P + tv \mid t \in \mathbb{R}\}$.

Example 45 $\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \middle| t \in \mathbb{R} \right\}$ describes a line in $\mathbb{R}^4$ parallel to the x_1-axis.

Given two non-zero vectors u, v, they will *usually* determine a plane,



unless both vectors are in the same line, in which case, one of the vectors is a scalar multiple of the other. The sum of u and v corresponds to laying the two vectors head-to-tail and drawing the connecting vector. If u and v determine a plane, then their sum lies in the plane determined by u and v.



The plane determined by two vectors u and v can be written as

$$\{P + su + tv \mid s, t \in \mathbb{R}\}\,.$$

Example 46 (A plane in a higher dimensional space)

$$\left\{ \begin{pmatrix} 3\\1\\4\\1\\5\\9 \end{pmatrix} + s \begin{pmatrix} 1\\0\\0\\0\\0\\0 \end{pmatrix} + t \begin{pmatrix} 0\\1\\0\\0\\0\\0 \end{pmatrix} \middle| \; s, t \in \mathbb{R} \right\}$$

describes a plane in 6-dimensional space parallel to the xy-plane.



`Parametric Notation`



We can generalize the notion of a plane with the following recursive definition. (That is, infinitely many things are defined in the following line.)

Definition A set of $k+1$ vectors $P, v_1, \ldots, v_k$ in $\mathbb{R}^n$ with $k \leq n$ determines a k-dimensional **hyperplane**,

$$\left\{ P + \sum_{i=1}^{k} \lambda_i v_i \,\middle|\, \lambda_i \in \mathbb{R} \right\}$$

unless any of the vectors v_j lives in the $(k-1)$-dimensional hyperplane determined by the other $k-1$ vectors

$$\left\{0+\sum_{i\neq j}^{k}\lambda_i v_i \,\middle|\, \lambda_i\in\mathbb{R}\right\}.$$

Example 47 (3+1 vectors that do not specify a 3-dimensional hyperplane)

$$S := \left\{ \begin{pmatrix}3\\1\\4\\1\\5\\9\end{pmatrix} + s\begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix} + t\begin{pmatrix}0\\1\\0\\0\\0\\0\end{pmatrix} + u\begin{pmatrix}1\\1\\0\\0\\0\\0\end{pmatrix} \,\middle|\, s,t,u\in\mathbb{R} \right\}$$

is not a 3-dimensional hyperplane because

$$\begin{pmatrix}1\\1\\0\\0\\0\\0\end{pmatrix} = 1\begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix} + 1\begin{pmatrix}0\\1\\0\\0\\0\\0\end{pmatrix} \in \left\{ s\begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix} + t\begin{pmatrix}0\\1\\0\\0\\0\\0\end{pmatrix} \,\middle|\, s,t\in\mathbb{R} \right\}.$$

In fact, the set could be rewritten as

$$S = \left\{ \begin{pmatrix}3\\1\\4\\1\\5\\9\end{pmatrix} + (s+u)\begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix} + (t+u)\begin{pmatrix}0\\1\\0\\0\\0\\0\end{pmatrix} \,\middle|\, s,t,u\in\mathbb{R} \right\}$$

$$= \left\{ \begin{pmatrix}3\\1\\4\\1\\5\\9\end{pmatrix} + a\begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix} + b\begin{pmatrix}0\\1\\0\\0\\0\\0\end{pmatrix} \,\middle|\, a,b\in\mathbb{R} \right\}$$

and so is actually the same 2-dimensional hyperplane in $\mathbb{R}^6$ as in example 46.

You might sometimes encounter the word "hyperplane" without the qualifier "k-dimensional. When the dimension k is not specified, one usually assumes that $k = n - 1$ for a hyperplane inside $\mathbb{R}^n$. This is the kind of object that is specified by one algebraic equation in n variables.

Example 48 (Specifying a plane with one linear algebraic equation.)
The solution set to

$$x_1 + x_2 + x_3 + x_4 + x_5 = 1 \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 1 - x_2 - x_3 - x_4 - x_5 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

is

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s_2 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + s_3 \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + s_4 \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + s_5 \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \middle| \, s_2, s_3, s_4, s_5 \in \mathbb{R} \right\},$$

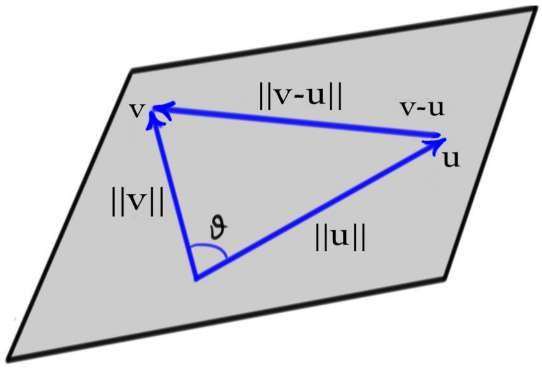
a 4-dimensional hyperplane in $\mathbb{R}^5$.

4.3 Directions and Magnitudes

Consider the *Euclidean length* of an n-vector:

$$\|v\| := \sqrt{(v^1)^2 + (v^2)^2 + \cdots + (v^n)^2} \;=\; \sqrt{\sum_{i=1}^{n} (v^i)^2}\,.$$

Using the Law of Cosines, we can then figure out the angle between two vectors. Given two vectors v and u that *span* a plane in $\mathbb{R}^n$, we can then connect the ends of v and u with the vector $v - u$.



Then the Law of Cosines states that:

$$\|v-u\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\,\|v\|\cos\theta$$

Then isolate $\cos\theta$:

$$\begin{aligned} \|v-u\|^2 - \|u\|^2 - \|v\|^2 &= (v^1-u^1)^2 + \cdots + (v^n-u^n)^2 \\ &\quad -\big((u^1)^2 + \cdots + (u^n)^2\big) \\ &\quad -\big((v^1)^2 + \cdots + (v^n)^2\big) \\ &= -2u^1v^1 - \cdots - 2u^nv^n \end{aligned}$$

Thus,

$$\|u\|\,\|v\|\cos\theta = u^1v^1 + \cdots + u^nv^n\,.$$

Note that in the above discussion, we have assumed (correctly) that Euclidean lengths in $\mathbb{R}^n$ give the usual notion of lengths of vectors for any plane in $\mathbb{R}^n$. This now motivates the definition of the dot product.

Definition The **dot product** of $u = \begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix}$ and $v = \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}$ is

$$u \cdot v := u^1v^1 + \cdots + u^nv^n\,.$$

Example 49 of the dot product of two vectors from $\mathbb{R}^{100}$.

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 100 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = 1 + 2 + 3 + \cdots + 100 = \frac{1}{2}.100.101 = 5050.$$

The sum above is the one Gauß, according to legend, could do in kindergarten.

Definition The **length** (or **norm** or **magnitude**) of an n-vector v is

$$\|v\| := \sqrt{v \cdot v}\,.$$

Example 50 of the norm of a vector from $\mathbb{R}^{101}$.

$$\left\| \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 101 \end{pmatrix} \right\| = \sqrt{\sum_{i=1}^{101} i^2} = \sqrt{37,961}.$$

Definition The **angle** θ between two vectors is determined by the formula

$$u \cdot v = \|u\|\|v\| \cos\theta\,.$$

Example 51 of an angle between two vectors form $\mathbb{R}^{101}$.

The angle between $\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ \vdots \\ 101 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$ is $\arccos\left(\frac{10,201}{\sqrt{37,916}\sqrt{51}}\right)$.

Definition Two vectors are **orthogonal** (or **perpendicular**) if their dot product is zero.

Example 52 of vectors from $\mathbb{R}^{101}$ that are orthogonal to each other.

$$\begin{pmatrix}1\\1\\1\\\vdots\\1\end{pmatrix}\cdot\begin{pmatrix}1\\-1\\1\\\vdots\\-1\end{pmatrix}=0.$$

Notice that the zero vector 0_n from $\mathbb{R}^n$ is orthogonal to every vector in $\mathbb{R}^n$; $0_n \cdot v = 0$ for all $v \in \mathbb{R}^n$.

The dot product has some important properties; it is

1. *symmetric*:
$$u \cdot v = v \cdot u\,,$$
2. *Distributive*:
$$u \cdot (v + w) = u \cdot v + u \cdot w\,,$$
3. *Bilinear* (which is to say, linear in both u and v):
$$u \cdot (cv + dw) = c\,u \cdot v + d\,u \cdot w\,,$$
and
$$(cu + dw) \cdot v = c\,u \cdot v + d\,w \cdot v\,.$$
4. *Positive Definite*:
$$u \cdot u \geq 0\,,$$
and $u \cdot u = 0$ only when u itself is the 0-vector.

There are, in fact, many different useful ways to define lengths of vectors. Notice in the definition above that we first defined the dot product, and then defined everything else in terms of the dot product. So if we change our idea of the dot product, we change our notion of length and angle as well. The dot product determines the *Euclidean length and angle* between two vectors.

Other definitions of length and angle arise from **inner products**, which have all of the properties listed above (except that in some contexts the positive definite requirement is relaxed). Instead of writing $\cdot$ for other inner products, we usually write $\langle u, v\rangle$ to avoid confusion.

Reading homework: problem 1

Example 53 Consider a four-dimensional space, with a special direction which we will call "time". The *Lorentzian inner product* on $\mathbb{R}^4$ is given by $\langle u, v\rangle = u^1v^1 + u^2v^2 + u^3v^3 - u^4v^4$. This is of central importance in Einstein's theory of special relativity. Note, in particular, that it is not positive definite. As a result, the "squared-length" of a vector with coordinates x, y, z and t is $\|v\|^2 = x^2 + y^2 + z^2 - t^2$. Notice that it is possible for $\|v\|^2 \leq 0$ even with non-vanishing v! The physical interpretation of this inner product depends on the sign of the inner product; two space time points $X_1 := (x_1, y_1, z_1, t_1),\ X_2 := (x_2, y_2, z_2, t_2)$ are

- separated by a distance $\sqrt{\langle X_1, X_2\rangle}$ if $\langle X_1, X_2\rangle \geq 0$.

- separated by a time $\sqrt{-\langle X_1, X_2\rangle}$ if $\langle X_1, X_2\rangle \leq 0$.

In particular, the difference in time coordinates $t_2 - t_1$ is not the time between the two points! (Compare this to using polar coordinates for which the distance between two points (r, θ_1) and (r, θ_2) is not $\theta_2 - \theta_1$; coordinate differences are not necessarily distances.)

Theorem 4.3.1 (Cauchy-Schwarz Inequality). *For any non-zero vectors u and v with an inner-product $\langle\ ,\ \rangle$*

$$\frac{|\langle u, v\rangle|}{\|u\|\,\|v\|} \leq 1.$$

The easiest proof would use the definition of the angle between two vectors and the fact that $\cos\theta \leq 1$. However, strictly speaking speaking we did not check our assumption that we could apply the Law of Cosines to the Euclidean length in $\mathbb{R}^n$. There is, however a simple algebraic proof.

Proof. Let α be any real number and consider the following positive, quadratic polynomial in α

$$0 \leq \langle u + \alpha v, u + \alpha v\rangle = \langle u, u\rangle + 2\alpha\langle u, v\rangle + \alpha^2\langle v, v\rangle\,.$$

Since any quadratic $a\alpha^2 + 2b\alpha + c$ takes its minimal value $c - \frac{b^2}{a}$ when $\alpha = -\frac{b}{2a}$, and the inequality should hold for even this minimum value of the polynomial

$$0 \leq \langle u, u\rangle - \frac{\langle u, v\rangle^2}{\langle v, v\rangle} \Leftrightarrow \frac{|\langle u, v\rangle|}{\|u\|\,\|v\|} \leq 1.$$

□

Theorem 4.3.2 (Triangle Inequality)**.** *For any* $u, v \in \mathbb{R}^n$

$$\|u+v\| \leq \|u\| + \|v\|.$$

Proof.

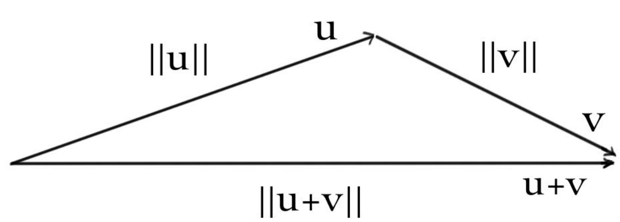
$$\begin{aligned} \|u+v\|^2 &= (u+v) \cdot (u+v) \\ &= u \cdot u + 2u \cdot v + v \cdot v \\ &= \|u\|^2 + \|v\|^2 + 2\,\|u\|\,\|v\| \cos\theta \\ &= (\|u\| + \|v\|)^2 + 2\,\|u\|\,\|v\|(\cos\theta - 1) \\ &\leq (\|u\| + \|v\|)^2 \,. \end{aligned}$$

That is, the square of the left-hand side of the triangle inequality is $\leq$ the square of the right-hand side. Since both the things being squared are positive, the inequality holds without the square;

$$\|u+v\| \leq \|u\| + \|v\|$$

□

The triangle inequality is also "self-evident" when examining a sketch of u, v and $u+v$.



Example 54 Let

$$a = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \text{ and } b = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix},$$

so that

$$a \cdot a = b \cdot b = 1 + 2^2 + 3^2 + 4^2 = 30$$

$$\Rightarrow \|a\| = \sqrt{30} = \|b\| \text{ and } \big(\|a\| + \|b\|\big)^2 = (2\sqrt{30})^2 = 120\,.$$

Since

$$a + b = \begin{pmatrix} 5 \\ 5 \\ 5 \\ 5 \end{pmatrix},$$

we have

$$\|a+b\|^2 = 5^2 + 5^2 + 5^2 + 5^2 = 100 < 120 = \big(\|a\| + \|b\|\big)^2$$

as predicted by the triangle inequality.

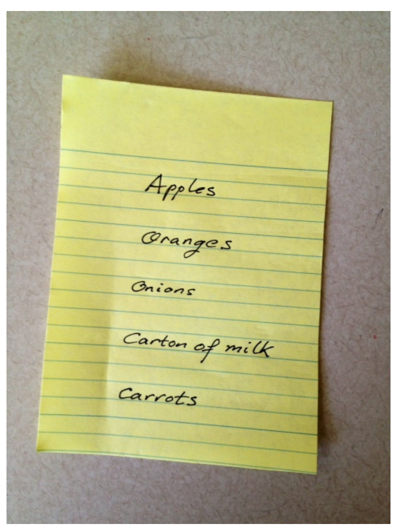
Notice also that $a \cdot b = 1.4 + 2.3 + 3.2 + 4.1 = 20 < \sqrt{30}.\sqrt{30} = 30 = \|a\|\,\|b\|$ in accordance with the Cauchy–Schwarz inequality.



Reading homework: problem 2

4.4 Vectors, Lists and Functions: $\mathbb{R}^S$

If you were going shopping you might make something like the following list.



We could represent this information mathematically as a set,

$$S = \{\text{apple}, \text{orange}, \text{onion}, \text{milk}, \text{carrot}\}\,.$$

There is no information of ordering here and no information about how many carrots you will buy. This set by itself is not a vector; how would we add such sets to one another?

If you were a more careful shopper your list might look like the following.

To buy at 5:23pm tonight:

onions	$\mapsto$	3.0
oranges	$\mapsto$	2.0
milk cartons	$\mapsto$	1.0
apples	$\mapsto$	5.0
carrots	$\mapsto$	8.0

What you have really done here is assign a number to each element of the set S. In other words, the second list is a function

$$f : S \longrightarrow \mathbb{R}\,.$$

Given two lists like the second one above, we could easily add them – if you plan to buy 5 apples and I am buying 3 apples, together we will buy 8 apples! In fact, the second list is really a 5-vector in disguise.

In general it is helpful to think of an n-vector as a function whose domain is the set $\{1, \ldots, n\}$. This is equivalent to thinking of an n-vector as an ordered list of n numbers. These two ideas give us two equivalent notions for the set of all n-vectors:

$$\mathbb{R}^n := \left\{ \begin{pmatrix} a^1 \\ \vdots \\ a^n \end{pmatrix} \middle| \, a^1, \ldots, a^n \in \mathbb{R} \right\} = \{a : \{1, \ldots, n\} \to \mathbb{R}\} =: \mathbb{R}^{\{1, \cdots, n\}}$$

The notation $\mathbb{R}^{\{1, \cdots, n\}}$ is used to denote the set of all functions from $\{1, \ldots, n\}$ to $\mathbb{R}$.

Similarly, for any set S the notation $\mathbb{R}^S$ denotes the set of functions from S to $\mathbb{R}$:

$$\mathbb{R}^S := \{f : S \to \mathbb{R}\}\,.$$

When S is an ordered set like $\{1, \ldots, n\}$, it is natural to write the components in order. When the elements of S do not have a natural ordering, doing so might cause confusion.

Example 55 Consider the set $S = \{*, \star, \#\}$ from chapter 1 review problem 9. A particular element of $\mathbb{R}^S$ is the function a explicitly defined by

$$a^\star = 3, a^\# = 5, a^* = -2.$$

It is not natural to write

$$a = \begin{pmatrix} 3 \\ 5 \\ -2 \end{pmatrix} \text{ or } a = \begin{pmatrix} -2 \\ 3 \\ 5 \end{pmatrix}$$

because the elements of S do not have an ordering, since as sets $\{*, \star, \#\} = \{\star, \#, *\}$.

In this important way, $\mathbb{R}^S$ seems different from $\mathbb{R}^3$. What is more evident are the similarities; since we can add two functions, we can add two elements of $\mathbb{R}^S$:

Example 56 Addition in $\mathbb{R}^{\{*, \star, \#\}}$
If $a, b \in \mathbb{R}^{\{*, \star, \#\}}$ such that

$$a^\star = 3, a^\# = 5, a^* = -2$$

and

$$b^\star = -2, b^\# = 4, b^* = 13$$

then $a + b \in \mathbb{R}^S$ is the function such that

$$(a+b)^\star = 3 - 2 = 1, (a+b)^\# = 5 + 4 = 9, (a+b)^* = -2 + 13 = 11\,.$$

Also, since we can multiply functions by numbers, there is a notion of scalar multiplication on $\mathbb{R}^S$.

Example 57 Scalar Multiplication in $\mathbb{R}^S$
If $a \in \mathbb{R}^{\{*, \star, \#\}}$ such that

$$a^\star = 3, a^\# = 5, a^* = -2$$

then $3a \in \mathbb{R}^{\{*, \star, \#\}}$ is the function such that

$$(3a)^\star = 3 \cdot 3 = 9, (3a)^\# = 3 \cdot 5 = 15, (3a)^* = 3(-2) = -6\,.$$

We visualize $\mathbb{R}^2$ and $\mathbb{R}^3$ in terms of axes. We have a more abstract picture of $\mathbb{R}^4$, $\mathbb{R}^5$ and $\mathbb{R}^n$ for larger n while $\mathbb{R}^S$ seems even more abstract. However, when thought of as a simple "shopping list", you can see that vectors in $\mathbb{R}^S$ in fact, can describe everyday objects. In chapter 5 we introduce the general definition of a vector space that unifies all these different notions of a vector.

4.5 Review Problems

Webwork:

Reading problems	1, 2
Vector operations	3
Vectors and lines	4
Vectors and planes	5
Lines, planes and vectors	6,7
Equation of a plane	8,9
Angle between a line and plane	10

1. When he was young, Captain Conundrum mowed lawns on weekends to help pay his college tuition bills. He charged his customers according to the size of their lawns at a rate of 5¢ per square foot and meticulously kept a record of the areas of their lawns in an ordered list:

$$A = (200, 300, 50, 50, 100, 100, 200, 500, 1000, 100)\,.$$

He also listed the number of times he mowed each lawn in a given year, for the year 1988 that ordered list was

$$f = (20, 1, 2, 4, 1, 5, 2, 1, 10, 6)\,.$$

(a) Pretend that A and f are vectors and compute $A \cdot f$.

(b) What quantity does the dot product $A \cdot f$ measure?

(c) How much did Captain Conundrum earn from mowing lawns in 1988? Write an expression for this amount in terms of the vectors A and f.

(d) Suppose Captain Conundrum charged different customers different rates. How could you modify the expression in part 1c to compute the Captain's earnings?

2. (2) Find the angle between the diagonal of the unit square in $\mathbb{R}^2$ and one of the coordinate axes.

(3) Find the angle between the diagonal of the unit cube in $\mathbb{R}^3$ and one of the coordinate axes.

(n) Find the angle between the diagonal of the unit (hyper)-cube in $\mathbb{R}^n$ and one of the coordinate axes.

(∞) What is the limit as $n \to \infty$ of the angle between the diagonal of the unit (hyper)-cube in $\mathbb{R}^n$ and one of the coordinate axes?

3. Consider the matrix $M = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ and the vector $X = \begin{pmatrix} x \\ y \end{pmatrix}$.

   (a) Sketch X and MX in $\mathbb{R}^2$ for several values of X and θ.

   (b) Compute $\frac{||MX||}{||X||}$ for arbitrary values of X and θ.

   (c) Explain your result for (b) and describe the action of M geometrically.

4. (Lorentzian Strangeness). For this problem, consider $\mathbb{R}^n$ with the Lorentzian inner product defined in example 53 above.

   (a) Find a non-zero vector in two-dimensional Lorentzian space-time with zero length.

   (b) Find and sketch the collection of all vectors in two-dimensional Lorentzian space-time with zero length.

   (c) Find and sketch the collection of all vectors in three-dimensional Lorentzian space-time with zero length.

   (d) Replace the word "zero" with the word "one" in the previous two parts.



The Story of Your Life



5. Create a system of equations whose solution set is a 99 dimensional hyperplane in $\mathbb{R}^{101}$.

6. Recall that a plane in $\mathbb{R}^3$ can be described by the equation

$$n \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = n \cdot p$$

where the vector p labels a given point on the plane and n is a vector normal to the plane. Let N and P be vectors in $\mathbb{R}^{101}$ and

$$X = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^{101} \end{pmatrix}.$$

What kind of geometric object does $N \cdot X = N \cdot P$ describe?

7. Let

$$u = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \text{ and } v = \begin{pmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ 101 \end{pmatrix}$$

Find the projection of v onto u and the projection of u onto v. (*Hint:* Remember that two vectors u and v define a plane, so first work out how to project one vector onto another in a plane. The picture from Section 14.4 could help.)

8. If the solution set to the equation $A(x) = b$ is the set of vectors whose tips lie on the paraboloid $z = x^2 + y^2$, then what can you say about the function A?

9. Find a system of equations whose solution set is

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} -1 \\ -1 \\ 0 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 0 \\ -1 \\ -3 \end{pmatrix} \middle| \, c_1, c_2 \in \mathbb{R} \right\}.$$

Give a general procedure for going from a parametric description of a hyperplane to a system of equations with that hyperplane as a solution set.

10. If A is a linear operator and both v and cv (for any real number c) are solutions to $Ax = b$, then what can you say about b?

5 Vector Spaces

As suggested at the end of chapter 4, the vector spaces $\mathbb{R}^n$ are not the only vector spaces. We now give a general definition that includes $\mathbb{R}^n$ for all values of n, and $\mathbb{R}^S$ for all sets S, and more. This mathematical structure is applicable to a wide range of real-world problems and allows for tremendous economy of thought; the idea of a basis for a vector space will drive home the main idea of vector spaces; they are sets with very simple structure.

The two key properties of vectors are that they can be added together and multiplied by scalars. Thus, before giving a rigorous definition of vector spaces, we restate the main idea.

> A vector space is a set that is closed under addition and scalar multiplication.

Definition A **vector space** $(V, +, . , \mathbb{R})$ is a set V with two operations $+$ and $\cdot$ satisfying the following properties for all $u, v \in V$ and $c, d \in \mathbb{R}$:

(+i) (Additive Closure) $u + v \in V$. *Adding two vectors gives a vector.*

(+ii) (Additive Commutativity) $u + v = v + u$. *Order of addition does not matter.*

(+iii) (Additive Associativity) $(u + v) + w = u + (v + w)$. *Order of adding many vectors does not matter.*

(+iv) (Zero) There is a special vector $0_V \in V$ such that $u + 0_V = u$ for all u in V.

(+v) (Additive Inverse) For every $u \in V$ there exists $w \in V$ such that $u + w = 0_V$.

($\cdot$ i) (Multiplicative Closure) $c \cdot v \in V$. *Scalar times a vector is a vector.*

($\cdot$ ii) (Distributivity) $(c+d) \cdot v = c \cdot v + d \cdot v$. *Scalar multiplication distributes over addition of scalars.*

($\cdot$ iii) (Distributivity) $c \cdot (u+v) = c \cdot u + c \cdot v$. *Scalar multiplication distributes over addition of vectors.*

($\cdot$ iv) (Associativity) $(cd) \cdot v = c \cdot (d \cdot v)$.

($\cdot$ v) (Unity) $1 \cdot v = v$ for all $v \in V$.

`Examples of each rule`

Remark Rather than writing $(V, +, ., \mathbb{R})$, we will often say "let V be a vector space over $\mathbb{R}$". If it is obvious that the numbers used are real numbers, then "let V be a vector space" suffices. Also, don't confuse the scalar product $\cdot$ with the dot product $\bullet$. The scalar product is a function that takes as its two inputs one number and one vector and returns a vector as its output. This can be written

$$\cdot \colon \mathbb{R} \times V \to V\,.$$

Similarly

$$+ : V \times V \to V\,.$$

On the other hand, the dot product takes two vectors and returns a number. Succinctly: $\bullet \colon V \times V \to \mathbb{R}$. Once the properties of a vector space have been verified, we'll just write scalar multiplication with juxtaposition $cv = c \cdot v$, though, to keep our notation efficient.

5.1 Examples of Vector Spaces

One can find many interesting vector spaces, such as the following:

Example of a vector space

Example 58

$$\mathbb{R}^{\mathbb{N}} = \{f \mid f\colon \mathbb{N} \rightarrow \mathbb{R}\}$$

Here the vector space is the set of functions that take in a natural number n and return a real number. The addition is just addition of functions: $(f_1+f_2)(n) = f_1(n)+f_2(n)$. Scalar multiplication is just as simple: $c \cdot f(n) = cf(n)$.

We can think of these functions as infinitely large ordered lists of numbers: $f(1) = 1^3 = 1$ is the first component, $f(2) = 2^3 = 8$ is the second, and so on. Then for example the function $f(n) = n^3$ would look like this:

$$f = \begin{pmatrix} 1 \\ 8 \\ 27 \\ \vdots \\ n^3 \\ \vdots \end{pmatrix}.$$

Thinking this way, $\mathbb{R}^{\mathbb{N}}$ is the space of all infinite sequences. Because we can not write a list infinitely long (without infinite time and ink), one can not define an element of this space explicitly; definitions that are implicit, as above, or algebraic as in $f(n) = n^3$ (for all $n \in \mathbb{N}$) suffice.

Let's check some axioms.

(+i) (Additive Closure) $(f_1 + f_2)(n) = f_1(n) + f_2(n)$ is indeed a function $\mathbb{N} \rightarrow \mathbb{R}$, since the sum of two real numbers is a real number.

(+iv) (Zero) We need to propose a zero vector. The constant zero function $g(n) = 0$ works because then $f(n) + g(n) = f(n) + 0 = f(n)$.

The other axioms should also be checked. This can be done using properties of the real numbers.

Reading homework: problem 1

Example 59 The space of functions of one real variable.

$$\mathbb{R}^{\mathbb{R}} = \{f \mid f\colon \mathbb{R} \rightarrow \mathbb{R}\}$$

The addition is point-wise

$$(f+g)(x) = f(x) + g(x)\,,$$

as is scalar multiplication

$$c \cdot f(x) = cf(x)\,.$$

To check that $\mathbb{R}^{\mathbb{R}}$ is a vector space use the properties of addition of functions and scalar multiplication of functions as in the previous example.

We can not write out an explicit definition for one of these functions either, there are not only infinitely many components, but even infinitely many components between any two components! You are familiar with algebraic definitions like $f(x) = e^{x^2-x+5}$. However, most vectors in this vector space can not be defined algebraically. For example, the nowhere continuous function

$$f(x) = \begin{cases} 1\,, & x \in \mathbb{Q} \\ 0\,, & x \notin \mathbb{Q} \end{cases}\,.$$

Example 60 $\mathbb{R}^{\{*,\star,\#\}} = \{f : \{*,\star,\#\} \to \mathbb{R}\}$. Again, the properties of addition and scalar multiplication of functions show that this is a vector space.

You can probably figure out how to show that $\mathbb{R}^S$ is vector space for any set S. This might lead you to guess that all vector spaces are of the form $\mathbb{R}^S$ for some set S. The following is a counterexample.

Example 61 Another very important example of a vector space is the space of all differentiable functions:

$$\left\{ f\colon \mathbb{R} \to \mathbb{R} \,\middle|\, \frac{d}{dx} f \text{ exists} \right\}.$$

From calculus, we know that the sum of any two differentiable functions is differentiable, since the derivative distributes over addition. A scalar multiple of a function is also differentiable, since the derivative commutes with scalar multiplication ($\frac{d}{dx}(cf) = c\frac{d}{dx}f$). The zero function is just the function such that $0(x) = 0$ for every x. The rest of the vector space properties are inherited from addition and scalar multiplication in $\mathbb{R}$.

Similarly, the set of functions with at least k derivatives is always a vector space, as is the space of functions with infinitely many derivatives. None of these examples can be written as $\mathbb{R}^S$ for some set S. Despite our emphasis on such examples, it is also not true that all vector spaces consist of functions. Examples are somewhat esoteric, so we omit them.

Another important class of examples is vector spaces that live inside $\mathbb{R}^n$ but are not themselves $\mathbb{R}^n$.

Example 62 (Solution set to a homogeneous linear equation.)
Let

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{pmatrix}.$$

The solution set to the homogeneous equation $Mx = 0$ is

$$\left\{ c_1 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \middle| c_1, c_2 \in \mathbb{R} \right\}.$$

This set is not equal to $\mathbb{R}^3$ since it does not contain, for example, $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. The sum of any two solutions is a solution, for example

$$\left[2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right] + \left[7 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + 5 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right] = 9 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + 8 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

and any scalar multiple of a solution is a solution

$$4 \left[5 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} - 3 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right] = 20 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} - 12 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

This example is called a *subspace* because it gives a vector space inside another vector space. See chapter 9 for details. Indeed, because it is determined by the linear map given by the matrix M, it is called $\ker M$, or in words, the *kernel* of M, for this see chapter 16.

Similarly, the solution set to any homogeneous linear equation is a vector space: Additive and multiplicative closure follow from the following statement, made using linearity of matrix multiplication:

If $Mx_1 = 0$ and $Mx_2 = 0$ then $M(c_1x_1 + c_2x_2) = c_1Mx_1 + c_2Mx_2 = 0 + 0 = 0$.

A powerful result, called the subspace theorem (see chapter 9) guarantees, based on the closure properties alone, that homogeneous solution sets are vector spaces.

More generally, if V is any vector space, then any hyperplane through the origin of V is a vector space.

Example 63 Consider the functions $f(x) = e^x$ and $g(x) = e^{2x}$ in $\mathbb{R}^\mathbb{R}$. By taking combinations of these two vectors we can form the plane $\{c_1 f + c_2 g | c_1, c_2 \in \mathbb{R}\}$ inside of $\mathbb{R}^\mathbb{R}$. This is a vector space; some examples of vectors in it are $4e^x - 31e^{2x}$, $\pi e^{2x} - 4e^x$ and $\frac{1}{2}e^{2x}$.

A hyperplane which does not contain the origin cannot be a vector space because it fails condition (+iv).

It is also possible to build new vector spaces from old ones using the product of sets. Remember that if V and W are sets, then their product is the new set

$$V \times W = \{(v, w) | v \in V, w \in W\},$$

or in words, all ordered pairs of elements from V and W. In fact $V \times W$ is a vector space if V and W are. We have actually been using this fact already:

Example 64 The real numbers $\mathbb{R}$ form a vector space (over $\mathbb{R}$). The new vector space

$$\mathbb{R} \times \mathbb{R} = \{(x, y) | x \in \mathbb{R}, y \in \mathbb{R}\}$$

has addition and scalar multiplication defined by

$$(x, y) + (x', y') = (x + x', y + y') \text{ and } c.(x, y) = (cx, cy).$$

Of course, this is just the vector space $\mathbb{R}^2 = \mathbb{R}^{\{1,2\}}$.

5.1.1 Non-Examples

The solution set to a linear non-homogeneous equation is not a vector space because it does not contain the zero vector and therefore fails (iv).

Example 65 The solution set to

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

is $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} -1 \\ 1 \end{pmatrix} \middle| c \in \mathbb{R} \right\}$. The vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is not in this set.

Do notice that if just one of the vector space rules is broken, the example is not a vector space.

Most sets of n-vectors are not vector spaces.

Example 66 $P := \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \middle| a, b \geq 0 \right\}$ is not a vector space because the set fails (·i) since $\begin{pmatrix} 1 \\ 1 \end{pmatrix} \in P$ but $-2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \end{pmatrix} \notin P$.

Sets of functions other than those of the form $\mathbb{R}^S$ should be carefully checked for compliance with the definition of a vector space.

Example 67 The set of all functions which are nowhere zero

$$\{f : \mathbb{R} \to \mathbb{R} \mid f(x) \neq 0 \text{ for any } x \in \mathbb{R}\} ,$$

does not form a vector space because it does not satisfy (+i). The functions $f(x) = x^2+1$ and $g(x) = -5$ are in the set, but their sum $(f+g)(x) = x^2-4 = (x+2)(x-2)$ is not since $(f + g)(2) = 0$.

5.2 Other Fields

Above, we defined vector spaces over the real numbers. One can actually define vector spaces over any *field*. This is referred to as choosing a different *base field*. A field is a collection of "numbers" satisfying properties which are listed in appendix B. An example of a field is the complex numbers,

$$\mathbb{C} = \left\{ x + iy \mid i^2 = -1, x, y \in \mathbb{R} \right\} .$$

Example 68 In quantum physics, vector spaces over $\mathbb{C}$ describe all possible states a physical system can have. For example,

$$V = \left\{ \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \mid \lambda, \mu \in \mathbb{C} \right\}$$

is the set of possible states for an electron's spin. The vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ describe, respectively, an electron with spin "up" and "down" along a given direction. Other vectors, like $\begin{pmatrix} i \\ -i \end{pmatrix}$ are permissible, since the base field is the complex numbers. Such states represent a mixture of spin up and spin down for the given direction (a rather counterintuitive yet experimentally verifiable concept), but a given spin in some other direction.

Complex numbers are very useful because of a special property that they enjoy: every polynomial over the complex numbers factors into a product of linear polynomials. For example, the polynomial

$$x^2 + 1$$

doesn't factor over real numbers, but over complex numbers it factors into

$$(x + i)(x - i)\,.$$

In other words, there are *two* solutions to

$$x^2 = -1,$$

$x = i$ and $x = -i$. This property has far-reaching consequences: often in mathematics problems that are very difficult using only real numbers become relatively simple when working over the complex numbers. This phenomenon occurs when diagonalizing matrices, see chapter 13.

The rational numbers $\mathbb{Q}$ are also a field. This field is important in computer algebra: a real number given by an infinite string of numbers after the decimal point can't be stored by a computer. So instead rational approximations are used. Since the rationals are a field, the mathematics of vector spaces still apply to this special case.

Another very useful field is bits

$$B_2 = \mathbb{Z}_2 = \{0, 1\}\,,$$

with the addition and multiplication rules

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

These rules can be summarized by the relation $2 = 0$. For bits, it follows that $-1 = 1$!

The theory of fields is typically covered in a class on abstract algebra or Galois theory.

5.3 Review Problems

Webwork:

Reading problems	1
Addition and inverse	2

1. Check that $\left\{\begin{pmatrix}x\\y\end{pmatrix}\middle| x,y\in\mathbb{R}\right\}=\mathbb{R}^2$ (with the usual addition and scalar multiplication) satisfies all of the parts in the definition of a vector space.

2. (a) Check that the complex numbers $\mathbb{C}=\{x+iy \mid i^2=-1, x,y\in\mathbb{R}\}$, satisfy all of the parts in the definition of a vector space over $\mathbb{C}$. Make sure you state carefully what your rules for vector addition and scalar multiplication are.

   (b) What would happen if you used $\mathbb{R}$ as the base field (try comparing to problem 1).

3. (a) Consider the set of convergent sequences, with the same addition and scalar multiplication that we defined for the space of sequences:
$$V=\left\{f \mid f\colon \mathbb{N}\to\mathbb{R}, \lim_{n\to\infty} f(n)\in\mathbb{R}\right\}\subset \mathbb{R}^{\mathbb{N}}.$$
   Is this still a vector space? Explain why or why not.

   (b) Now consider the set of divergent sequences, with the same addition and scalar multiplication as before:
$$V=\left\{f \mid f\colon \mathbb{N}\to\mathbb{R}, \lim_{n\to\infty} f(n) \text{ does not exist or is } \pm\infty\right\}\subset \mathbb{R}^{\mathbb{N}}.$$
   Is this a vector space? Explain why or why not.

4. Consider the set of 2×4 matrices:
$$V=\left\{\begin{pmatrix}a & b & c & d\\ e & f & g & h\end{pmatrix}\middle| a,b,c,d,e,f,g,h\in\mathbb{C}\right\}$$

Propose definitions for addition and scalar multiplication in V. Identify the zero vector in V, and check that every matrix in V has an additive inverse.

5. Let $P_3^{\mathbb{R}}$ be the set of polynomials with real coefficients of degree three or less.

   (a) Propose a definition of addition and scalar multiplication to make $P_3^{\mathbb{R}}$ a vector space.

   (b) Identify the zero vector, and find the additive inverse for the vector $-3 - 2x + x^2$.

   (c) Show that $P_3^{\mathbb{R}}$ is not a vector space over $\mathbb{C}$. Propose a small change to the definition of $P_3^{\mathbb{R}}$ to make it a vector space over $\mathbb{C}$. (Hint: Every little symbol in the the instructions for par (c) is important!)



Hint



6. Let $V = \{x \in \mathbb{R} | x > 0\} =: \mathbb{R}_+$. For $x, y \in V$ and $\lambda \in \mathbb{R}$, define

$$x \oplus y = xy\,, \qquad \lambda \otimes x = x^\lambda\,.$$

   Show that $(V, \oplus, \otimes, \mathbb{R})$ is a vector space.

7. The component in the ith row and jth column of a matrix can be labeled m^i_j. In this sense a matrix is a function of a pair of integers. For what set S is the set of 2×2 matrices the same as the set $\mathbb{R}^S$? Generalize to other size matrices.

8. Show that any function in $\mathbb{R}^{\{*,\star,\#\}}$ can be written as a sum of multiples of the functions $e_*, e_\star, e_\#$ defined by

$$e_*(k) = \begin{cases} 1\,, & k = * \\ 0\,, & k = \star \\ 0\,, & k = \# \end{cases}\,,\ e_\star(k) = \begin{cases} 0\,, & k = * \\ 1\,, & k = \star \\ 0\,, & k = \# \end{cases}\,,\ e_\#(k) = \begin{cases} 0\,, & k = * \\ 0\,, & k = \star \\ 1\,, & k = \# \end{cases}\,.$$

9. Let V be a vector space and S any set. Show that the set V^S of all functions $S \to V$ is a vector space. *Hint:* first decide upon a rule for adding functions whose outputs are vectors.

6

Linear Transformations

The main objects of study in any course in linear algebra are linear functions:

Definition A function $L\colon V \to W$ is **linear** if V and W are vector spaces and

$$L(ru + sv) = rL(u) + sL(v)$$

for all $u, v \in V$ and $r, s \in \mathbb{R}$.

Reading homework: problem 1

Remark We will often refer to linear functions by names like "linear map", "linear operator" or "linear transformation". In some contexts you will also see the name "homomorphism" which generally is applied to functions from one kind of set to the same kind of set while respecting any structures on the sets; linear maps are from vector spaces to vector spaces that respect scalar multiplication and addition, the two structures on vector spaces. It is common to denote a linear function by capital L as a reminder of its linearity, but sometimes we will use just f, after all we are just studying very special functions.

The definition above coincides with the two part description in Chapter 1; the case $r = 1, s = 1$ describes additivity, while $s = 0$ describes homogeneity. We are now ready to learn the powerful consequences of linearity.

6.1 The Consequence of Linearity

Now that we have a sufficiently general notion of vector space it is time to talk about why linear operators are so special. Think about what is required to fully specify a real function of one variable. One output must be specified for each input. That is an infinite amount of information.

By contrast, even though a linear function can have infinitely many elements in its domain, it is specified by a very small amount of information.

Example 69 (One output specifies infinitely many)
If you know that the function L is linear and that

$$L\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}5\\3\end{pmatrix}$$

then you do not need any more information to figure out

$$L\begin{pmatrix}2\\0\end{pmatrix},\ L\begin{pmatrix}3\\0\end{pmatrix}\ ,L\begin{pmatrix}4\\0\end{pmatrix},\ L\begin{pmatrix}5\\0\end{pmatrix},\ etc\ldots,$$

because by homogeneity

$$L\begin{pmatrix}5\\0\end{pmatrix}=L\left[5\begin{pmatrix}1\\0\end{pmatrix}\right]=5L\begin{pmatrix}1\\0\end{pmatrix}=5\begin{pmatrix}5\\3\end{pmatrix}=\begin{pmatrix}25\\15\end{pmatrix}.$$

In this way an an infinite number of outputs is specified by just one.

Example 70 (Two outputs in $\mathbb{R}^2$ specifies all outputs)
Likewise, if you know that L is linear and that

$$L\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}5\\3\end{pmatrix}\text{ and }L\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}2\\2\end{pmatrix}$$

then you don't need any more information to compute

$$L\begin{pmatrix}1\\1\end{pmatrix}$$

because by additivity

$$L\begin{pmatrix}1\\1\end{pmatrix}=L\left[\begin{pmatrix}1\\0\end{pmatrix}+\begin{pmatrix}0\\1\end{pmatrix}\right]=L\begin{pmatrix}1\\0\end{pmatrix}+L\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}5\\3\end{pmatrix}+\begin{pmatrix}2\\2\end{pmatrix}=\begin{pmatrix}7\\5\end{pmatrix}.$$

In fact, since every vector in $\mathbb{R}^2$ can be expressed as

$$\begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} ,$$

we know how L acts on every vector from $\mathbb{R}^2$ by linearity based on just two pieces of information;

$$L \begin{pmatrix} x \\ y \end{pmatrix} = L \left[x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right] = xL \begin{pmatrix} 1 \\ 0 \end{pmatrix} + yL \begin{pmatrix} 0 \\ 1 \end{pmatrix} = x \begin{pmatrix} 5 \\ 3 \end{pmatrix} + y \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 5x + 2y \\ 3x + 2y \end{pmatrix} .$$

Thus, the value of L at infinitely many inputs is completely specified by its value at just two inputs. (We can see now that L acts in exactly the way the matrix

$$\begin{pmatrix} 5 & 2 \\ 3 & 2 \end{pmatrix}$$

acts on vectors from $\mathbb{R}^2$.)

Reading homework: problem 2

This is the reason that linear functions are so nice; they are secretly very simple functions by virtue of two characteristics:

1. They act on vector spaces.

2. They act additively and homogeneously.

A linear transformation with domain $\mathbb{R}^3$ is completely specified by the way it acts on the three vectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} , \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} , \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} .$$

Similarly, a linear transformation with domain $\mathbb{R}^n$ is completely specified by its action on the n different n-vectors that have exactly one non-zero component, and its matrix form can be read off this information. However, not all linear functions have such nice domains.

6.2 Linear Functions on Hyperplanes

It is not always so easy to write a linear operator as a matrix. Generally, this will amount to solving a linear systems problem. Examining a linear function whose domain is a hyperplane is instructive.

Example 71 Let

$$V = \left\{ c_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \middle| c_1, c_2 \in \mathbb{R} \right\}$$

and consider $L : V \to \mathbb{R}^3$ be a linear function that obeys

$$L \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad L \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

By linearity this specifies the action of L on any vector from V as

$$L \left[c_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right] = (c_1 + c_2) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

The domain of L is a plane and its range is the line through the origin in the x_2 direction.

It is not clear how to formulate L as a matrix; since

$$L \begin{pmatrix} c_1 \\ c_1 + c_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_1 + c_2 \\ c_2 \end{pmatrix} = (c_1 + c_2) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

or

$$L \begin{pmatrix} c_1 \\ c_1 + c_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_1 + c_2 \\ c_2 \end{pmatrix} = (c_1 + c_2) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},$$

you might suspect that L is equivalent to one of these 3×3 matrices. It is not. By the natural domain convention, all 3×3 matrices have $\mathbb{R}^3$ as their domain, and the domain of L is smaller than that. When we do realize this L as a matrix it will be as a 3×2 matrix. We can tell because the domain of L is 2 dimensional and the codomain is 3 dimensional. (You probably already know that the plane has dimension 2, and a

line is 1 dimensional, but the careful definition of "dimension" takes some work; this is tackled in Chapter 11.) This leads us to write

$$L\left[c_1\begin{pmatrix}1\\1\\0\end{pmatrix}+c_2\begin{pmatrix}0\\1\\1\end{pmatrix}\right]=c_1\begin{pmatrix}0\\1\\0\end{pmatrix}+c_2\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}0&0\\1&1\\0&0\end{pmatrix}\begin{pmatrix}c_1\\c_2\end{pmatrix}.$$

This makes sense, but requires a *warning*: The matrix $\begin{pmatrix}0&0\\1&1\\0&0\end{pmatrix}$ specifies L so long as you also provide the information that you are labeling points in the plane V by the two numbers (c_1, c_2).

6.3 Linear Differential Operators

Your calculus class became much easier when you stopped using the limit definition of the derivative, learned the power rule, and started using linearity of the derivative operator.

Example 72 Let V be the vector space of polynomials of degree 2 or less with standard addition and scalar multiplication;

$$V := \{a_0 \cdot 1 + a_1 x + a_2 x^2 \,|\, a_0, a_1, a_2 \in \mathbb{R}\}$$

Let $\frac{d}{dx} : V \to V$ be the derivative operator. The following three equations, along with linearity of the derivative operator, allow one to take the derivative of any 2nd degree polynomial:

$$\frac{d}{dx}1 = 0,\ \frac{d}{dx}x = 1,\ \frac{d}{dx}x^2 = 2x\,.$$

In particular

$$\frac{d}{dx}(a_0 1 + a_1 x + a_2 x^2) = a_0\frac{d}{dx}1 + a_1\frac{d}{dx}x + a_2\frac{d}{dx}x^2 = 0 + a_1 + 2a_2 x.$$

Thus, the derivative acting any of the infinitely many second order polynomials is determined by its action for just three inputs.

6.4 Bases (Take 1)

The central idea of linear algebra is to exploit the hidden simplicity of linear functions. It ends up there is a lot of freedom in how to do this. That freedom is what makes linear algebra powerful.

You saw that a linear operator acting on $\mathbb{R}^2$ is completely specified by how it acts on the pair of vectors $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$. In fact, any linear operator acting on $\mathbb{R}^2$ is also completely specified by how it acts on the pair of vectors $\begin{pmatrix}1\\1\end{pmatrix}$ and $\begin{pmatrix}1\\-1\end{pmatrix}$.

Example 73 The linear operator L is a linear operator then it is completely specified by the two equalities

$$L\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}2\\4\end{pmatrix},\ \text{and}\ L\begin{pmatrix}1\\-1\end{pmatrix}=\begin{pmatrix}6\\8\end{pmatrix}.$$

This is because any vector $\begin{pmatrix}x\\y\end{pmatrix}$ in $\mathbb{R}^2$ is a sum of multiples of $\begin{pmatrix}1\\1\end{pmatrix}$ and $\begin{pmatrix}1\\-1\end{pmatrix}$ which can be calculated via a linear systems problem as follows:

$$\begin{aligned}
&\begin{pmatrix}x\\y\end{pmatrix}=a\begin{pmatrix}1\\1\end{pmatrix}+b\begin{pmatrix}1\\-1\end{pmatrix}\\
\Leftrightarrow\ &\begin{pmatrix}1&1\\1&-1\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}=\begin{pmatrix}x\\y\end{pmatrix}\\
\Leftrightarrow\ &\left(\begin{array}{rr|r}1&1&x\\1&-1&y\end{array}\right)\sim\left(\begin{array}{rr|r}1&0&\frac{x+y}{2}\\0&1&\frac{x-y}{2}\end{array}\right)\\
\Leftrightarrow\ &\left\{\begin{array}{l}a=\frac{x+y}{2}\\b=\frac{x-y}{2}\end{array}\right..
\end{aligned}$$

Thus

$$\begin{pmatrix}x\\y\end{pmatrix}=\frac{x+y}{2}\begin{pmatrix}1\\1\end{pmatrix}+\frac{x-y}{2}\begin{pmatrix}1\\-1\end{pmatrix}.$$

We can then calculate how L acts on any vector by first expressing the vector as a

sum of multiples and then applying linearity;

$$\begin{aligned}
L\begin{pmatrix}x\\y\end{pmatrix} &= L\left[\frac{x+y}{2}\begin{pmatrix}1\\1\end{pmatrix}+\frac{x-y}{2}\begin{pmatrix}1\\-1\end{pmatrix}\right]\\
&= \frac{x+y}{2}L\begin{pmatrix}1\\1\end{pmatrix}+\frac{x-y}{2}L\begin{pmatrix}1\\-1\end{pmatrix}\\
&= \frac{x+y}{2}\begin{pmatrix}2\\4\end{pmatrix}+\frac{x-y}{2}\begin{pmatrix}6\\8\end{pmatrix}\\
&= \begin{pmatrix}x+y\\2(x+y)\end{pmatrix}+\begin{pmatrix}3(x-y)\\4(x-y)\end{pmatrix}\\
&= \begin{pmatrix}4x-2y\\6x-y\end{pmatrix}
\end{aligned}$$

Thus L is completely specified by its value at just two inputs.

It should not surprise you to learn there are infinitely many pairs of vectors from $\mathbb{R}^2$ with the property that any vector can be expressed as a linear combination of them; any pair that when used as columns of a matrix gives an invertible matrix works. Such a pair is called a *basis* for $\mathbb{R}^2$.

Similarly, there are infinitely many triples of vectors with the property that any vector from $\mathbb{R}^3$ can be expressed as a linear combination of them: these are the triples that used as columns of a matrix give an invertible matrix. Such a triple is called a basis for $\mathbb{R}^3$.

In a similar spirit, there are infinitely many pairs of vectors with the property that every vector in

$$V=\left\{c_1\begin{pmatrix}1\\1\\0\end{pmatrix}+c_2\begin{pmatrix}0\\1\\1\end{pmatrix}\middle|\, c_1,c_2\in\mathbb{R}\right\}$$

can be expressed as a linear combination of them. Some examples are

$$V=\left\{c_1\begin{pmatrix}1\\1\\0\end{pmatrix}+c_2\begin{pmatrix}0\\2\\2\end{pmatrix}\middle|\, c_1,c_2\in\mathbb{R}\right\}=\left\{c_1\begin{pmatrix}1\\1\\0\end{pmatrix}+c_2\begin{pmatrix}1\\3\\2\end{pmatrix}\middle|\, c_1,c_2\in\mathbb{R}\right\}$$

Such a pair is a called a basis for V.

You probably have some intuitive notion of what dimension means (the careful mathematical definition is given in chapter 11). Roughly speaking,

dimension is the number of independent directions available. To figure out the dimension of a vector space, I stand at the origin, and pick a direction. If there are any vectors in my vector space that aren't in that direction, then I choose another direction that isn't in the line determined by the direction I chose. If there are any vectors in my vector space not in the plane determined by the first two directions, then I choose one of them as my next direction. In other words, I choose a collection of *independent* vectors in the vector space (independent vectors are defined in Chapter 10). A minimal set of independent vectors is called a *basis* (see Chapter 11 for the precise definition). The number of vectors in my basis is the dimension of the vector space. Every vector space has many bases, but all bases for a particular vector space have the same number of vectors. Thus dimension is a well-defined concept.

The fact that every vector space (over $\mathbb{R}$) has infinitely many bases is actually very useful. Often a good choice of basis can reduce the time required to run a calculation in dramatic ways!

In summary:

> A basis is a set of vectors in terms of which it is possible to uniquely express any other vector.

6.5 Review Problems

Webwork:

Reading problems	1, 2
Linear?	3
Matrix $\times$ vector	4, 5
Linearity	6, 7

1. Show that the pair of conditions:

$$\left\{ \begin{array}{l} L(u+v) = L(u) + L(v) \\ L(cv) = cL(v) \end{array} \right. \tag{1}$$

(valid for all vectors u, v and any scalar c) is equivalent to the single condition:

$$L(ru + sv) = rL(u) + sL(v)\,, \tag{2}$$

(for all vectors u, v and any scalars r and s). Your answer should have two parts. Show that (1) $\Rightarrow$ (2), and then show that (2) $\Rightarrow$ (1).

2. If f is a linear function of one variable, then how many points on the graph of the function are needed to specify the function? Give an explicit expression for f in terms of these points. (You might want to look up the definition of a graph before you make any assumptions about the function.)

3. (a) If $p\begin{pmatrix}1\\2\end{pmatrix} = 1$ and $p\begin{pmatrix}2\\4\end{pmatrix} = 3$ is it possible that p is a linear function?
   (b) If $Q(x^2) = x^3$ and $Q(2x^2) = x^4$ is it possible that Q is a linear function from polynomials to polynomials?

4. If f is a linear function such that

$$f\begin{pmatrix}1\\2\end{pmatrix} = 0, \text{ and } f\begin{pmatrix}2\\3\end{pmatrix} = 1\,,$$

   then what is $f\begin{pmatrix}x\\y\end{pmatrix}$?

5. Let P_n be the space of polynomials of degree n or less in the variable t. Suppose L is a linear transformation from $P_2 \to P_3$ such that $L(1) = 4$, $L(t) = t^3$, and $L(t^2) = t - 1$.
   (a) Find $L(1 + t + 2t^2)$.
   (b) Find $L(a + bt + ct^2)$.
   (c) Find all values a, b, c such that $L(a + bt + ct^2) = 1 + 3t + 2t^3$.

Hint

6. Show that the operator $\mathcal{I}$ that maps f to the function $\mathcal{I}f$ defined by $\mathcal{I}f(x) := \int_0^x f(t)dt$ is a linear operator on the space of continuous functions.

7. Let $z \in \mathbb{C}$. Recall that $z = x+iy$ for some $x, y \in \mathbb{R}$, and we can form the *complex conjugate* of z by taking $\overline{z} = x - iy$. The function $c\colon \mathbb{R}^2 \to \mathbb{R}^2$ which sends $(x, y) \mapsto (x, -y)$ agrees with complex conjugation.
   (a) Show that c is a linear map over $\mathbb{R}$ (*i.e.* scalars in $\mathbb{R}$).
   (b) Show that $\overline{z}$ is not linear over $\mathbb{C}$.

7 Matrices

Matrices are a powerful tool for calculations involving linear transformations. It is important to understand how to find the matrix of a linear transformation and the properties of matrices.

7.1 Linear Transformations and Matrices

Ordered, finite-dimensional, bases for vector spaces allows us to express linear operators as matrices.

7.1.1 Basis Notation

A basis allows us to efficiently label arbitrary vectors in terms of column vectors. Here is an example.

Example 74 Let

$$V = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| a, b, c, d \in \mathbb{R} \right\}$$

be the vector space of 2×2 real matrices, with addition and scalar multiplication defined componentwise. One choice of basis is the ordered set (or list) of matrices

$$B = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right) =: (e^1_1, e^1_2, e^2_1, e^2_2)\,.$$

Given a particular vector and a basis, your job is to write that vector as a sum of multiples of basis elements. Here an arbitrary vector $v \in V$ is just a matrix, so we write

$$\begin{aligned} v = \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ c & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & d \end{pmatrix} \\ &= a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\ &= a\, e_1^1 + b\, e_2^1 + c\, e_1^2 + d\, e_2^2 . \end{aligned}$$

The coefficients (a, b, c, d) of the basis vectors $(e_1^1, e_2^1, e_1^2, e_2^2)$ encode the information of which matrix the vector v is. We store them in column vector by writing

$$v = a\, e_1^1 + b\, e_2^1 + c\, e_1^2 + d\, e_2^2 =: (e_1^1, e_2^1, e_1^2, e_2^2) \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} =: \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}_B .$$

The 4-vector $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \in \mathbb{R}^4$ encodes the vector $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in V$ but is NOT equal to it! (After all, v is a matrix so could not equal a column vector.) Both notations on the right hand side of the above equation really stand for the vector obtained by multiplying the coefficients stored in the column vector by the corresponding basis element and then summing over them.

Next, lets consider a tautological example showing how to label column vectors in terms of column vectors:

Example 75 (Standard Basis of $\mathbb{R}^2$)
The vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

are called the standard basis vectors of $\mathbb{R}^2 = \mathbb{R}^{\{1,2\}}$. Their description as functions of $\{1, 2\}$ are

$$e_1(k) = \begin{cases} 1 & \text{if } k = 1 \\ 0 & \text{if } k = 2 \end{cases}, \quad e_2(k) = \begin{cases} 0 & \text{if } k = 1 \\ 1 & \text{if } k = 2 . \end{cases}$$

It is natural to assign these the order: e_1 is first and e_2 is second. An arbitrary vector v of $\mathbb{R}^2$ can be written as

$$v = \begin{pmatrix} x \\ y \end{pmatrix} = xe_1 + ye_2.$$

To emphasize that we are using the standard basis we define the list (or ordered set)

$$E = (e_1, e_2)\,,$$

and write

$$\begin{pmatrix} x \\ y \end{pmatrix}_E := (e_1, e_2)\begin{pmatrix} x \\ y \end{pmatrix} := xe_1 + ye_2 = v.$$

You should read this equation by saying:

"The column vector of the vector v in the basis E is $\begin{pmatrix} x \\ y \end{pmatrix}$."

Again, the first notation of a column vector with a subscript E refers to the vector obtained by multiplying each basis vector by the corresponding scalar listed in the column and then summing these, $i.e.$ $xe_1 + ye_2$. The second notation denotes exactly the same thing but we first list the basis elements and then the column vector; a useful trick because this can be read in the same way as matrix multiplication of a row vector times a column vector–except that the entries of the row vector are themselves vectors!

You should already try to write down the standard basis vectors for $\mathbb{R}^n$ for other values of n and express an arbitrary vector in $\mathbb{R}^n$ in terms of them.

The last example probably seems pedantic because column vectors are already just ordered lists of numbers and the basis notation has simply allowed us to "re-express" these as lists of numbers. Of course, this objection does not apply to more complicated vector spaces like our first matrix example. Moreover, as we saw earlier, there are infinitely many other pairs of vectors in $\mathbb{R}^2$ that form a basis.

Example 76 (A Non-Standard Basis of $\mathbb{R}^2 = \mathbb{R}^{\{1,2\}}$)

$$b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\,, \quad \beta = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

As functions of $\{1, 2\}$ they read

$$b(k) = \begin{cases} 1 & \text{if } k = 1 \\ 1 & \text{if } k = 2 \end{cases}\,, \quad \beta(k) = \begin{cases} 1 & \text{if } k = 1 \\ -1 & \text{if } k = 2\,. \end{cases}$$

Notice something important: there is no reason to say that β comes before b or vice versa. That is, there is no *a priori* reason to give these basis elements one order or the other. However, it will be necessary to give the basis elements an order if we want to use them to encode other vectors. We choose one arbitrarily; let

$$B = (b, \beta)$$

be the ordered basis. Note that for an unordered set we use the $\{\}$ parentheses while for lists or ordered sets we use $()$.

As before we define

$$\begin{pmatrix} x \\ y \end{pmatrix}_B := (b, \beta) \begin{pmatrix} x \\ y \end{pmatrix} := xb + y\beta \,.$$

You might think that the numbers x and y denote exactly the same vector as in the previous example. However, they do not. Inserting the actual vectors that b and β represent we have

$$xb + y\beta = x \begin{pmatrix} 1 \\ 1 \end{pmatrix} + y \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} x+y \\ x-y \end{pmatrix} \,.$$

Thus, to contrast, we have

$$\begin{pmatrix} x \\ y \end{pmatrix}_B = \begin{pmatrix} x+y \\ x-y \end{pmatrix} \text{ and } \begin{pmatrix} x \\ y \end{pmatrix}_E = \begin{pmatrix} x \\ y \end{pmatrix}$$

Only in the standard basis E does the column vector of v agree with the column vector that v actually is!

Based on the above example, you might think that our aim would be to find the "standard basis" for any problem. In fact, this is far from the truth. Notice, for example that the vector

$$v = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = e_1 + e_2 = b$$

written in the standard basis E is just

$$v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}_E \,,$$

which was easy to calculate. But in the basis B we find

$$v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}_B \,,$$

which is actually a simpler column vector! The fact that there are many bases for any given vector space allows us to choose a basis in which our computation is easiest. In any case, the standard basis only makes sense for $\mathbb{R}^n$. Suppose your vector space was the set of solutions to a differential equation–what would a standard basis then be?

Example 77 (A Basis For a Hyperplane)
Lets again consider the hyperplane

$$V = \left\{ c_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \middle| c_1, c_2 \in \mathbb{R} \right\}$$

One possible choice of ordered basis is

$$b_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad B = (b_1, b_2).$$

With this choice

$$\begin{pmatrix} x \\ y \end{pmatrix}_B := x b_1 + y b_2 = x \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ x + y \\ y \end{pmatrix}_E .$$

With the other choice of order $B' = (b_2, b_1)$

$$\begin{pmatrix} x \\ y \end{pmatrix}_{B'} := x b_2 + y b_1 = x \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + y \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} y \\ x + y \\ x \end{pmatrix}_E .$$

We see that the order of basis elements matters.

Finding the column vector of a given vector in a given basis usually amounts to a linear systems problem:

Example 78 (Pauli Matrices)
Let

$$V = \left\{ \begin{pmatrix} z & u \\ v & -z \end{pmatrix} \middle| z, u, v \in \mathbb{C} \right\}$$

be the vector space of trace-free complex-valued matrices (over $\mathbb{C}$) with basis

$$B = (\sigma_x, \sigma_y, \sigma_z),$$

where

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

These three matrices are the famous *Pauli matrices*; they are used to describe electrons in quantum theory, or qubits in quantum computation. Let

$$v = \begin{pmatrix} -2+i & 1+i \\ 3-i & 2-i \end{pmatrix}.$$

Find the column vector of v in the basis B.

For this we must solve the equation

$$\begin{pmatrix} -2+i & 1+i \\ 3-i & 2-i \end{pmatrix} = \alpha^x \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \alpha^y \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + \alpha^z \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

This gives four equations, *i.e.* a linear systems problem, for the α's

$$\left\{ \begin{array}{rrrcr} \alpha^x & -\, i\alpha^y & & = & 1+i \\ \alpha^x & +\, i\alpha^y & & = & 3-i \\ & & \alpha^z & = & -2+i \\ & & -\alpha^z & = & 2-i \end{array} \right.$$

with solution

$$\alpha^x = 2, \quad \alpha^y = 2-2i, \quad \alpha^z = -2+i.$$

Thus

$$v = \begin{pmatrix} -2+i & 1+i \\ 3-i & 2-i \end{pmatrix} = \begin{pmatrix} 2 \\ 2-i \\ -2+i \end{pmatrix}_B.$$

To summarize, the *column vector of a vector* v in an ordered basis $B = (b_1, b_2, \ldots, b_n)$,

$$\begin{pmatrix} \alpha^1 \\ \alpha^2 \\ \vdots \\ \alpha^n \end{pmatrix},$$

is defined by solving the linear systems problem

$$v = \alpha^1 b_1 + \alpha^2 b_2 + \cdots + \alpha^n b_n = \sum_{i=1}^{n} \alpha^i b_i.$$

The numbers $(\alpha^1, \alpha^2, \ldots, \alpha^n)$ are called the *components of the vector* v. Two useful shorthand notations for this are

$$v = \begin{pmatrix} \alpha^1 \\ \alpha^2 \\ \vdots \\ \alpha^n \end{pmatrix}_B = (b_1, b_2, \ldots, b_n) \begin{pmatrix} \alpha^1 \\ \alpha^2 \\ \vdots \\ \alpha^n \end{pmatrix} .$$

7.1.2 From Linear Operators to Matrices

Chapter 6 showed that linear functions are very special kinds of functions; they are fully specified by their values on any basis for their domain. A matrix records how a linear operator maps an element of the basis to a sum of multiples in the target space basis.

More carefully, if L is a linear operator from V to W then the matrix for L in the ordered bases $B = (b_1, b_2, \ldots)$ for V and $B' = (\beta_1, \beta_2, \ldots)$ for W, is the array of numbers m_i^j specified by

$$L(b_i) = m_i^1 \beta_1 + \cdots + m_i^j \beta_j + \cdots$$

Remark To calculate the matrix of a linear transformation you must compute what the linear transformation does to every input basis vector and then write the answers in terms of the output basis vectors:

$\big(L(b_1), L(b_2), \ldots, L(b_j), \ldots\big)$

$$= \left((\beta_1, \beta_2, \ldots, \beta_j, \ldots) \begin{pmatrix} m_1^1 \\ m_2^2 \\ \vdots \\ m_1^j \\ \vdots \end{pmatrix}, (\beta_1, \beta_2, \ldots, \beta_j, \ldots) \begin{pmatrix} m_2^1 \\ m_2^2 \\ \vdots \\ m_2^j \\ \vdots \end{pmatrix}, \cdots, (\beta_1, \beta_2, \ldots, \beta_j, \ldots) \begin{pmatrix} m_i^1 \\ m_i^2 \\ \vdots \\ m_i^j \\ \vdots \end{pmatrix}, \cdots \right)$$

$$= (\beta_1, \beta_2, \ldots, \beta_j, \ldots) \begin{pmatrix} m_1^1 & m_2^1 & \cdots & m_i^1 & \cdots \\ m_1^2 & m_2^2 & \cdots & m_i^2 & \cdots \\ \vdots & \vdots & & \vdots & \\ m_1^j & m_2^j & \cdots & m_i^j & \cdots \\ \vdots & \vdots & & \vdots & \end{pmatrix}$$

Example 79 Consider $L : V \to \mathbb{R}^3$ (as in example 71) defined by

$$L \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} , \quad L \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} .$$

By linearity this specifies the action of L on any vector from V as

$$L\left[c_1\begin{pmatrix}1\\1\\0\end{pmatrix}+c_2\begin{pmatrix}0\\1\\1\end{pmatrix}\right]=(c_1+c_2)\begin{pmatrix}0\\1\\0\end{pmatrix}.$$

We had trouble expressing this linear operator as a matrix. Lets take input basis

$$B=\left(\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}0\\1\\1\end{pmatrix}\right)=:(b_1,b_2)\,,$$

and output basis

$$E=\left(\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}\right).$$

Then

$$Lb_1=0e_1+1e_2+0e_3\,,$$
$$Lb_2=0e_1+1e_2+0e_3\,,$$

or

$$\big(Lb_1,Lb_2\big)=\big((e_1,e_2,e_3)\begin{pmatrix}0\\1\\0\end{pmatrix},(e_1,e_2,e_3)\begin{pmatrix}0\\1\\0\end{pmatrix}\big)=(e_1,e_2,e_3)\begin{pmatrix}0&0\\1&1\\0&0\end{pmatrix}.$$

The matrix on the right is the matrix of L in these bases. More succinctly we could write

$$L\begin{pmatrix}x\\y\end{pmatrix}_B=(x+y)\begin{pmatrix}0\\1\\0\end{pmatrix}_E$$

and thus see that L acts like the matrix $\begin{pmatrix}0&0\\1&1\\0&0\end{pmatrix}$.

Hence

$$L\begin{pmatrix}x\\y\end{pmatrix}_B=\left(\begin{pmatrix}0&0\\1&1\\0&0\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}\right)_E;$$

given input and output bases, the linear operator is now encoded by a matrix.

This is the general rule for this chapter:

Linear operators become matrices when given
ordered input and output bases.

Reading homework: problem 1

Example 80 Lets compute a matrix for the derivative operator acting on the vector space of polynomials of degree 2 or less:

$$V = \{a_0 1 + a_1 x + a_2 x^2 \,|\, a_0, a_1, a_2 \in \mathbb{R}\}\,.$$

In the ordered basis $B = (1, x, x^2)$ we write

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}_B = a \cdot 1 + bx + cx^2$$

and

$$\frac{d}{dx}\begin{pmatrix} a \\ b \\ c \end{pmatrix}_B = b \cdot 1 + 2cx + 0x^2 = \begin{pmatrix} b \\ 2c \\ 0 \end{pmatrix}_B$$

In the ordered basis B for both domain and range

$$\frac{d}{dx} \stackrel{B}{\mapsto} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

Notice this last line makes no sense without explaining which bases we are using!

7.2 Review Problems

Webwork:

Reading problem	1
Matrix of a Linear Transformation	9, 10, 11, 12, 13

1. A door factory can buy supplies in two kinds of packages, f and g. The package f contains 3 slabs of wood, 4 fasteners, and 6 brackets. The package g contains 5 fasteners, 3 brackets, and 7 slabs of wood.

   (a) Explain how to view the packages f and g as functions and list their inputs and outputs.

(b) Choose an ordering for the 3 kinds of supplies and use this to rewrite f and g as elements of $\mathbb{R}^3$.

(c) Let L be a manufacturing process that takes as inputs supply packages and outputs two products (doors, and door frames). Explain how it can be viewed as a function mapping one vector space into another.

(d) Assuming that L is linear and Lf is 1 door and 2 frames, and Lg is 3 doors and 1 frame, find a matrix for L. Be sure to specify the basis vectors you used, both for the input and output vector space.

2. You are designing a simple keyboard synthesizer with two keys. If you push the first key with intensity a then the speaker moves in time as $a\sin(t)$. If you push the second key with intensity b then the speaker moves in time as $b\sin(2t)$. If the keys are pressed simultaneously,

   (a) describe the set of all sounds that come out of your synthesizer. (*Hint:* Sounds can be "added".)

   (b) Graph the function $\begin{pmatrix}3\\5\end{pmatrix} \in \mathbb{R}^{\{1,2\}}$.

   (c) Let $B = (\sin(t), \sin(2t))$. Explain why $\begin{pmatrix}3\\5\end{pmatrix}_B$ is not in $\mathbb{R}^{\{1,2\}}$ but is still a function.

   (d) Graph the function $\begin{pmatrix}3\\5\end{pmatrix}_B$.

3. (a) Find the matrix for $\frac{d}{dx}$ acting on the vector space V of polynomials of degree 2 or less in the ordered basis $B = (x^2, x, 1)$

   (b) Use the matrix from part (a) to rewrite the differential equation $\frac{d}{dx}p(x) = x$ as a matrix equation. Find all solutions of the matrix equation. Translate them into elements of V.

(c) Find the matrix for $\frac{d}{dx}$ acting on the vector space V in the ordered basis $B' = (x^2 + x, x^2 - x, 1)$.

(d) Use the matrix from part (c) to rewrite the differential equation $\frac{d}{dx}p(x) = x$ as a matrix equation. Find all solutions of the matrix equation. Translate them into elements of V.

(e) Compare and contrast your results from parts (b) and (d).

4. Find the "matrix" for $\frac{d}{dx}$ acting on the vector space of all power series in the ordered basis $(1, x, x^2, x^3, ...)$. Use this matrix to find all power series solutions to the differential equation $\frac{d}{dx}f(x) = x$. *Hint:* your "matrix" may not have finite size.

5. Find the matrix for $\frac{d^2}{dx^2}$ acting on $\{c_1 \cos(x) + c_2 \sin(x) \mid c_1, c_2 \in \mathbb{R}\}$ in the ordered basis $(\cos(x), \sin(x))$.

6. Find the matrix for $\frac{d}{dx}$ acting on $\{c_1 \cosh(x) + c_2 \sinh(x) | c_1, c_2 \in \mathbb{R}\}$ in the ordered basis

$$(\cosh(x), \sinh(x))$$

and in the ordered basis

$$(\cosh(x) + \sinh(x), \cosh(x) - \sinh(x)).$$

7. Let $B = (1, x, x^2)$ be an ordered basis for

$$V = \{a_0 + a_1 x + a_2 x^2 \mid a_0, a_1, a_2 \in \mathbb{R}\},$$

and let $B' = (x^3, x^2, x, 1)$ be an ordered basis for

$$W = \{a_0 + a_1 x + a_2 x^2 + a_3 x^3 \mid a_0, a_1, a_2, a_3 \in \mathbb{R}\},$$

Find the matrix for the operator $\mathcal{I} : V \to W$ defined by

$$\mathcal{I}p(x) = \int_1^x p(t)dt$$

relative to these bases.

8. This exercise is meant to show you a generalization of the procedure you learned long ago for finding the function $mx+b$ given two points on its graph. It will also show you a way to think of matrices as members of a much bigger class of arrays of numbers.

   Find the

   (a) constant function $f : \mathbb{R} \to \mathbb{R}$ whose graph contains $(2, 3)$.

   (b) linear function $h : \mathbb{R} \to \mathbb{R}$ whose graph contains $(5, 4)$.

   (c) first order polynomial function $g : \mathbb{R} \to \mathbb{R}$ whose graph contains $(1, 2)$ and $(3, 3)$.

   (d) second order polynomial function $p : \mathbb{R} \to \mathbb{R}$ whose graph contains $(1, 0)$, $(3, 0)$ and $(5, 0)$.

   (e) second order polynomial function $q : \mathbb{R} \to \mathbb{R}$ whose graph contains $(1, 1)$, $(3, 2)$ and $(5, 7)$.

   (f) second order homogeneous polynomial function $r : \mathbb{R} \to \mathbb{R}$ whose graph contains $(3, 2)$.

   (g) number of points required to specify a third order polynomial $\mathbb{R} \to \mathbb{R}$.

   (h) number of points required to specify a third order homogeneous polynomial $\mathbb{R} \to \mathbb{R}$.

   (i) number of points required to specify a n-th order polynomial $\mathbb{R} \to \mathbb{R}$.

   (j) number of points required to specify a n-th order homogeneous polynomial $\mathbb{R} \to \mathbb{R}$.

   (k) first order polynomial function $F : \mathbb{R}^2 \to \mathbb{R}$ whose graph contains $\left(\begin{pmatrix}0\\0\end{pmatrix}, 1\right)$, $\left(\begin{pmatrix}0\\1\end{pmatrix}, 2\right)$, $\left(\begin{pmatrix}1\\0\end{pmatrix}, 3\right)$, and $\left(\begin{pmatrix}1\\1\end{pmatrix}, 4\right)$.

   (l) homogeneous first order polynomial function $H : \mathbb{R}^2 \to \mathbb{R}$ whose graph contains $\left(\begin{pmatrix}0\\1\end{pmatrix}, 2\right)$, $\left(\begin{pmatrix}1\\0\end{pmatrix}, 3\right)$, and $\left(\begin{pmatrix}1\\1\end{pmatrix}, 4\right)$.

(m) second order polynomial function $J : \mathbb{R}^2 \to \mathbb{R}$ whose graph contains $\left(\begin{pmatrix}0\\0\end{pmatrix}, 0\right)$, $\left(\begin{pmatrix}0\\1\end{pmatrix}, 2\right)$, $\left(\begin{pmatrix}0\\2\end{pmatrix}, 5\right)$, $\left(\begin{pmatrix}1\\0\end{pmatrix}, 3\right)$, $\left(\begin{pmatrix}2\\0\end{pmatrix}, 6\right)$, and $\left(\begin{pmatrix}1\\1\end{pmatrix}, 4\right)$.

(n) first order polynomial function $K : \mathbb{R}^2 \to \mathbb{R}^2$ whose graph contains $\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right)$, $\left(\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}2\\2\end{pmatrix}\right)$, $\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}3\\3\end{pmatrix}\right)$, and $\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}4\\4\end{pmatrix}\right)$.

(o) How many points in the graph of a q-th order polynomial function $\mathbb{R}^n \to \mathbb{R}^n$ would completely determine the function?

(p) In particular, how many points of the graph of linear function $\mathbb{R}^n \to \mathbb{R}^n$ would completely determine the function? How does a matrix (in the standard basis) encode this information?

(q) Propose a way to store the information required in 8g above in an array of numbers.

(r) Propose a way to store the information required in 8o above in an array of numbers.

7.3 Properties of Matrices

The objects of study in linear algebra are linear operators. We have seen that linear operators can be represented as matrices through choices of ordered bases, and that matrices provide a means of efficient computation.

We now begin an in depth study of matrices.

Definition An $r \times k$ **matrix** $M = (m^i_j)$ for $i = 1, \ldots, r; j = 1, \ldots, k$ is a rectangular array of real (or complex) numbers:

$$M = \begin{pmatrix} m^1_1 & m^1_2 & \cdots & m^1_k \\ m^2_1 & m^2_2 & \cdots & m^2_k \\ \vdots & \vdots & & \vdots \\ m^r_1 & m^r_2 & \cdots & m^r_k \end{pmatrix}.$$

The numbers m^i_j are called **entries**. The superscript indexes the row of the matrix and the subscript indexes the column of the matrix in which m^i_j appears.

An $r \times 1$ matrix $v = (v^r_1) = (v^r)$ is called a *column vector*, written

$$v = \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^r \end{pmatrix} .$$

A $1 \times k$ matrix $v = (v^1_k) = (v_k)$ is called a *row vector*, written

$$v = \begin{pmatrix} v_1 & v_2 & \cdots & v_k \end{pmatrix} .$$

The *transpose of a column vector* is the corresponding row vector and vice versa:

Example 81 Let

$$v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} .$$

Then

$$v^T = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} ,$$

and $(v^T)^T = v$. This is an example of an *involution*, namely an operation which when performed twice does nothing.

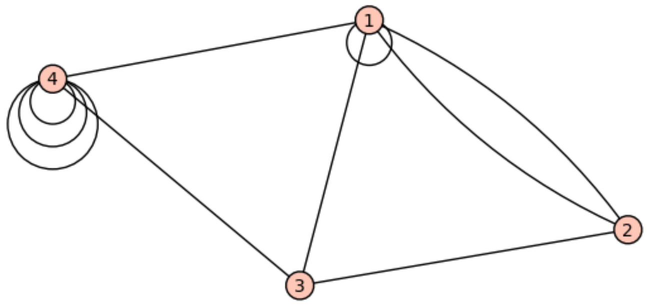
A matrix is an efficient way to store information.

Example 82 In computer graphics, you may have encountered image files with a .gif extension. These files are actually just matrices: at the start of the file the size of the matrix is given, after which each number is a matrix entry indicating the color of a particular pixel in the image.

This matrix then has its rows shuffled a bit: by listing, say, every eighth row, a web browser downloading the file can start displaying an incomplete version of the picture before the download is complete.

Finally, a compression algorithm is applied to the matrix to reduce the file size.

Example 83 Graphs occur in many applications, ranging from telephone networks to airline routes. In the subject of *graph theory*, a graph is just a collection of vertices and some edges connecting vertices. A matrix can be used to indicate how many edges attach one vertex to another.



For example, the graph pictured above would have the following matrix, where m^i_j indicates the number of edges between the vertices labeled i and j:

$$M = \begin{pmatrix} 1 & 2 & 1 & 1 \\ 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 3 \end{pmatrix}$$

This is an example of a *symmetric matrix*, since $m^i_j = m^j_i$.



`Adjacency Matrix Example`



The set of all $r \times k$ matrices

$$\mathbb{M}^r_k := \{(m^i_j) | m^i_j \in \mathbb{R};\ i \in \{1, \ldots, r\};\ j \in \{1 \ldots k\}\},$$

is itself a vector space with addition and scalar multiplication defined as follows:

$$M + N = (m^i_j) + (n^i_j) = (m^i_j + n^i_j)$$
$$rM = r(m^i_j) = (rm^i_j)$$

In other words, addition just adds corresponding entries in two matrices, and scalar multiplication multiplies every entry. Notice that $M_1^n = \mathbb{R}^n$ is just the vector space of column vectors.

Recall that we can multiply an $r \times k$ matrix by a $k \times 1$ column vector to produce a $r \times 1$ column vector using the rule

$$MV = \big(\sum_{j=1}^{k} m_j^i v^j\big) .$$

This suggests the rule for multiplying an $r \times k$ matrix M by a $k \times s$ matrix N: our $k \times s$ matrix N consists of s column vectors side-by-side, each of dimension $k \times 1$. We can multiply our $r \times k$ matrix M by each of these s column vectors using the rule we already know, obtaining s column vectors each of dimension $r \times 1$. If we place these s column vectors side-by-side, we obtain an $r \times s$ matrix MN.

That is, let

$$N = \begin{pmatrix} n_1^1 & n_2^1 & \cdots & n_s^1 \\ n_1^2 & n_2^2 & \cdots & n_s^2 \\ \vdots & \vdots & & \vdots \\ n_1^k & n_2^k & \cdots & n_s^k \end{pmatrix}$$

and call the columns N_1 through N_s:

$$N_1 = \begin{pmatrix} n_1^1 \\ n_1^2 \\ \vdots \\ n_1^k \end{pmatrix}, \; N_2 = \begin{pmatrix} n_2^1 \\ n_2^2 \\ \vdots \\ n_2^k \end{pmatrix}, \; \ldots, \; N_s = \begin{pmatrix} n_s^1 \\ n_s^2 \\ \vdots \\ n_s^k \end{pmatrix}.$$

Then

$$MN = M \begin{pmatrix} | & | & & | \\ N_1 & N_2 & \cdots & N_s \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ MN_1 & MN_2 & \cdots & MN_s \\ | & | & & | \end{pmatrix}$$

Concisely: If $M = (m_j^i)$ for $i = 1, \ldots, r; j = 1, \ldots, k$ and $N = (n_j^i)$ for $i = 1, \ldots, k; j = 1, \ldots, s$, then $MN = L$ where $L = (\ell_j^i)$ for $i = i, \ldots, r; j = 1, \ldots, s$ is given by

$$\ell_j^i = \sum_{p=1}^{k} m_p^i n_j^p.$$

This rule obeys linearity.

Notice that in order for the multiplication to make sense, the columns and rows must match. For an $r \times k$ matrix M and an $s \times m$ matrix N, then to make the product MN we must have $k = s$. Likewise, for the product NM, it is required that $m = r$. A common shorthand for keeping track of the sizes of the matrices involved in a given product is the following diagram.

$$\big(r \times k\big) \text{ times } \big(k \times m\big) \text{ is } \big(r \times m\big)$$

Reading homework: problem 2

Example 84 Multiplying a (3×1) matrix and a (1×2) matrix yields a (3×2) matrix.

$$\begin{pmatrix}1\\3\\2\end{pmatrix}\begin{pmatrix}2 & 3\end{pmatrix} = \begin{pmatrix}1\cdot2 & 1\cdot3\\3\cdot2 & 3\cdot3\\2\cdot2 & 2\cdot3\end{pmatrix} = \begin{pmatrix}2 & 3\\6 & 9\\4 & 6\end{pmatrix}.$$

Another way to view matrix multiplication is in terms of dot products:

The entries of MN are made from the dot products of the rows of M with the columns of N.

Example 85 Let

$$M = \begin{pmatrix}1 & 3\\3 & 5\\2 & 6\end{pmatrix} =: \begin{pmatrix}u^T\\v^T\\w^T\end{pmatrix} \text{ and } N = \begin{pmatrix}2 & 3 & 1\\0 & 1 & 0\end{pmatrix} =: \begin{pmatrix}a & b & c\end{pmatrix}$$

where

$$u = \begin{pmatrix}1\\3\end{pmatrix}, \quad v = \begin{pmatrix}3\\5\end{pmatrix}, \quad w = \begin{pmatrix}2\\6\end{pmatrix}, \quad a = \begin{pmatrix}2\\0\end{pmatrix}, \quad b = \begin{pmatrix}3\\1\end{pmatrix}, \quad c = \begin{pmatrix}1\\0\end{pmatrix}.$$

Then

$$MN = \begin{pmatrix}u\cdot a & u\cdot b & u\cdot c\\v\cdot a & v\cdot b & v\cdot c\\w\cdot a & w\cdot b & w\cdot c\end{pmatrix} = \begin{pmatrix}2 & 6 & 1\\6 & 14 & 3\\4 & 12 & 2\end{pmatrix}.$$

This fact has an obvious yet important consequence:

Theorem 7.3.1. *Let M be a matrix and x a column vector. If*

$$Mx = 0$$

then the vector x is orthogonal to the rows of M.

Remark Remember that the set of all vectors that can be obtained by adding up scalar multiples of the columns of a matrix is called its column space . Similarly the **row space** is the set of all row vectors obtained by adding up multiples of the rows of a matrix. The above theorem says that if $Mx = 0$, then the vector x is orthogonal to every vector in the row space of M.

We know that $r \times k$ matrices can be used to represent linear transformations $\mathbb{R}^k \to \mathbb{R}^r$ via

$$(MV)^i = \sum_{j=1}^{k} m^i_j v^j ,$$

which is the same rule used when we multiply an $r \times k$ matrix by a $k \times 1$ vector to produce an $r \times 1$ vector.

Likewise, we can use a matrix $N = (n^i_j)$ to define a linear transformation of a vector space of matrices. For example

$$L \colon M^s_k \stackrel{N}{\longrightarrow} M^r_k ,$$

$$L(M) = (l^i_k) \text{ where } l^i_k = \sum_{j=1}^{s} n^i_j m^j_k .$$

This is the same as the rule we use to multiply matrices. In other words, $L(M) = NM$ is a linear transformation.

Matrix Terminology Let $M = (m^i_j)$ be a matrix. The entries m^i_i are called **diagonal**, and the set $\{m^1_1, m^2_2, \ldots\}$ is called the **diagonal** of the matrix.

Any $r \times r$ matrix is called a **square matrix**. A square matrix that is zero for all non-diagonal entries is called a **diagonal matrix**. An example of a square diagonal matrix is

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix} .$$

The $r \times r$ diagonal matrix with all diagonal entries equal to 1 is called the **identity matrix**, I_r, or just I. An identity matrix looks like

$$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

The identity matrix is special because

$$I_r M = M I_k = M$$

for all M of size $r \times k$.

Definition The **transpose** of an $r \times k$ matrix $M = (m^i_j)$ is the $k \times r$ matrix

$$M^T = (\hat{m}^i_j)$$

with entries that satisfy $\hat{m}^i_j = m^j_i$.

A matrix M is *symmetric* if $M = M^T$.

Example 86

$$\begin{pmatrix} 2 & 5 & 6 \\ 1 & 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 2 & 1 \\ 5 & 3 \\ 6 & 4 \end{pmatrix},$$

and

$$\begin{pmatrix} 2 & 5 & 6 \\ 1 & 3 & 4 \end{pmatrix} \begin{pmatrix} 2 & 5 & 6 \\ 1 & 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 65 & 43 \\ 43 & 26 \end{pmatrix},$$

is symmetric.



Reading homework: problem 3

Observations

- Only square matrices can be symmetric.
- The transpose of a column vector is a row vector, and vice-versa.

- Taking the transpose of a matrix twice does nothing. *i.e.*, $(M^T)^T = M$.

Theorem 7.3.2 (Transpose and Multiplication)**.** *Let M, N be matrices such that MN makes sense. Then*

$$(MN)^T = N^T M^T .$$

The proof of this theorem is left to Review Question 2.

7.3.1 Associativity and Non-Commutativity

Many properties of matrices following from the same property for real numbers. Here is an example.

Example 87 *Associativity of matrix multiplication.* We know for real numbers x, y and z that

$$x(yz) = (xy)z\,,$$

$i.e.$, the order of multiplications does not matter. The same property holds for matrix multiplication, let us show why. Suppose $M = \big(m^i_j\big)$, $N = \big(n^j_k\big)$ and $R = \big(r^k_l\big)$ are, respectively, $m \times n$, $n \times r$ and $r \times t$ matrices. Then from the rule for matrix multiplication we have

$$MN = \Big(\sum_{j=1}^{n} m^i_j n^j_k\Big) \text{ and } NR = \Big(\sum_{k=1}^{r} n^j_k r^k_l\Big)\,.$$

So first we compute

$$(MN)R = \Big(\sum_{k=1}^{r}\Big[\sum_{j=1}^{n} m^i_j n^j_k\Big] r^k_l\Big) = \Big(\sum_{k=1}^{r}\sum_{j=1}^{n}\Big[m^i_j n^j_k\Big] r^k_l\Big) = \Big(\sum_{k=1}^{r}\sum_{j=1}^{n} m^i_j n^j_k r^k_l\Big)\,.$$

In the first step we just wrote out the definition for matrix multiplication, in the second step we moved summation symbol outside the bracket (this is just the distributive property $x(y+z) = xy+xz$ for numbers) and in the last step we used the associativity property for real numbers to remove the square brackets. Exactly the same reasoning shows that

$$M(NR) = \Big(\sum_{j=1}^{n} m^i_j\Big[\sum_{k=1}^{r} n^j_k r^k_l\Big]\Big) = \Big(\sum_{k=1}^{r}\sum_{j=1}^{n} m^i_j\Big[n^j_k r^k_l\Big]\Big) = \Big(\sum_{k=1}^{r}\sum_{j=1}^{n} m^i_j n^j_k r^k_l\Big)\,.$$

This is the same as above so we are done. [1]

[1] As a fun remark, note that Einstein would simply have written $(MN)R = (m^i_j n^j_k) r^k_l = m^i_j n^j_k r^k_l = m^i_j (n^j_k r^k_l) = M(NR)$.

Sometimes matrices do not share the properties of regular numbers. In particular, for *generic* $n \times n$ square matrices M and N,

$$\boxed{MN \neq NM\,.}$$



Do Matrices Commute?



Example 88 (Matrix multiplication does *not* commute.)

$$\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix} = \begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}$$

while, on the other hand,

$$\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix} = \begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}.$$

Since $n \times n$ matrices are linear transformations $\mathbb{R}^n \to \mathbb{R}^n$, we can see that the order of successive linear transformations matters.

Here is an example of matrices acting on objects in three dimensions that also shows matrices not commuting.

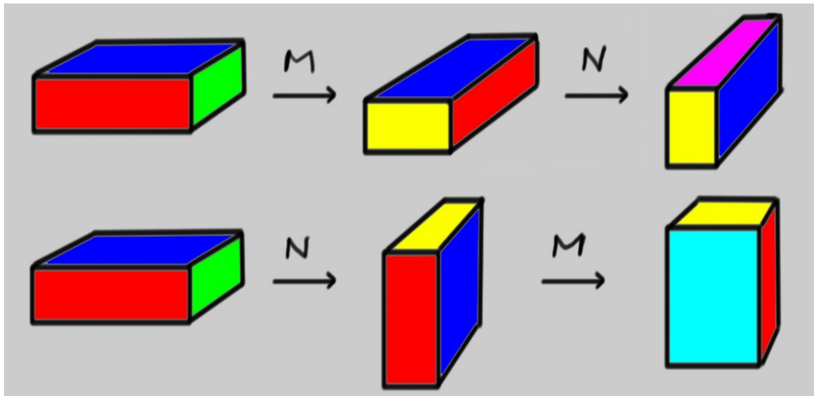
Example 89 In Review Problem 3, you learned that the matrix

$$M = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix},$$

rotates vectors in the plane by an angle θ. We can generalize this, using block matrices, to three dimensions. In fact the following matrices built from a 2×2 rotation matrix, a 1×1 identity matrix and zeroes everywhere else

$$M = \begin{pmatrix}\cos\theta & \sin\theta & 0\\ -\sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix} \qquad \text{and} \qquad N = \begin{pmatrix}1 & 0 & 0\\ 0 & \cos\theta & \sin\theta\\ 0 & -\sin\theta & \cos\theta\end{pmatrix},$$

perform rotations by an angle θ in the xy and yz planes, respectively. Because, they rotate single vectors, you can also use them to rotate objects built from a collection of vectors like pretty colored blocks! Here is a picture of M and then N acting on such a block, compared with the case of N followed by M. The special case of $\theta = 90^\circ$ is shown.



Notice how the endproducts of MN and NM are different, so $MN \neq NM$ here.

7.3.2 Block Matrices

It is often convenient to partition a matrix M into smaller matrices called *blocks*. For example

$$M = \left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 1 \\ \hline 0 & 1 & 2 & 0 \end{array}\right) = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$$

Where $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$, $B = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $C = \begin{pmatrix} 0 & 1 & 2 \end{pmatrix}$, $D = (0)$.

- The blocks of a block matrix must fit together to form a rectangle. So $\left(\begin{array}{c|c} B & A \\ \hline D & C \end{array}\right)$ makes sense, but $\left(\begin{array}{c|c} C & B \\ \hline D & A \end{array}\right)$ does not.

 Reading homework: problem 4

- There are many ways to cut up an $n \times n$ matrix into blocks. Often context or the entries of the matrix will suggest a useful way to divide the matrix into blocks. For example, if there are large blocks of zeros in a matrix, or blocks that look like an identity matrix, it can be useful to partition the matrix accordingly.

- Matrix operations on block matrices can be carried out by treating the blocks as matrix entries. In the example above,

$$\begin{aligned} M^2 &= \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \\ &= \left(\begin{array}{c|c} A^2+BC & AB+BD \\ \hline CA+DC & CB+D^2 \end{array}\right) \end{aligned}$$

Computing the individual blocks, we get:

$$\begin{aligned} A^2+BC &= \begin{pmatrix} 30 & 37 & 44 \\ 66 & 81 & 96 \\ 102 & 127 & 152 \end{pmatrix} \\ AB+BD &= \begin{pmatrix} 4 \\ 10 \\ 16 \end{pmatrix} \\ CA+DC &= \begin{pmatrix} 4 & 10 & 16 \end{pmatrix} \\ CB+D^2 &= (2) \end{aligned}$$

Assembling these pieces into a block matrix gives:

$$\left(\begin{array}{ccc|c} 30 & 37 & 44 & 4 \\ 66 & 81 & 96 & 10 \\ 102 & 127 & 152 & 16 \\ \hline 4 & 10 & 16 & 2 \end{array}\right)$$

This is exactly M^2.

7.3.3 The Algebra of Square Matrices

Not every pair of matrices can be multiplied. When multiplying two matrices, the number of rows in the left matrix must equal the number of columns in the right. For an $r \times k$ matrix M and an $s \times l$ matrix N, then we must have $k = s$.

This is not a problem for square matrices of the same size, though. Two $n \times n$ matrices can be multiplied in either order. For a single matrix $M \in \mathbb{M}_n^n$, we can form $M^2 = MM$, $M^3 = MMM$, and so on. It is useful

to define

$$M^0 = I\,,$$

the identity matrix, just like $x^0 = 1$ for numbers.

As a result, any polynomial can be have square matrices in it's domain.

Example 90 Let $f(x) = x - 2x^2 + 3x^3$ and

$$M = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Then

$$M^2 = \begin{pmatrix} 1 & 2t \\ 0 & 1 \end{pmatrix}\,,\ M^3 = \begin{pmatrix} 1 & 3t \\ 0 & 1 \end{pmatrix}\,,\ \ldots$$

and so

$$\begin{aligned} f(M) &= \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} - 2\begin{pmatrix} 1 & 2t \\ 0 & 1 \end{pmatrix} + 3\begin{pmatrix} 1 & 3t \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 6t \\ 0 & 2 \end{pmatrix}. \end{aligned}$$

Suppose $f(x)$ is any function defined by a convergent Taylor Series:

$$f(x) = f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \cdots\,.$$

Then we can define the matrix function by just plugging in M:

$$f(M) = f(0) + f'(0)M + \frac{1}{2!}f''(0)M^2 + \cdots\,.$$

There are additional techniques to determine the convergence of Taylor Series of matrices, based on the fact that the convergence problem is simple for diagonal matrices. It also turns out that the matrix exponential

$$\exp(M) = I + M + \frac{1}{2}M^2 + \frac{1}{3!}M^3 + \cdots\,,$$

always converges.

`Matrix Exponential Example`

7.3.4 Trace

A large matrix contains a great deal of information, some of which often reflects the fact that you have not set up your problem efficiently. For example, a clever choice of basis can often make the matrix of a linear transformation very simple. Therefore, finding ways to extract the essential information of a matrix is useful. Here we need to assume that $n < \infty$ otherwise there are subtleties with convergence that we'd have to address.

Definition The **trace** of a square matrix $M = (m^i_j)$ is the sum of its diagonal entries:

$$\operatorname{tr} M = \sum_{i=1}^{n} m^i_i .$$

Example 91

$$\operatorname{tr} \begin{pmatrix} 2 & 7 & 6 \\ 9 & 5 & 1 \\ 4 & 3 & 8 \end{pmatrix} = 2 + 5 + 8 = 15 .$$

While matrix multiplication does not commute, the trace of a product of matrices does not depend on the order of multiplication:

$$\begin{aligned} \operatorname{tr}(MN) &= \operatorname{tr}\Big(\sum_l M^i_l N^l_j\Big) \\ &= \sum_i \sum_l M^i_l N^l_i \\ &= \sum_l \sum_i N^l_i M^i_l \\ &= \operatorname{tr}\Big(\sum_i N^l_i M^i_l\Big) \\ &= \operatorname{tr}(NM). \end{aligned}$$



Proof Explanation



Thus we have a Theorem:

Theorem 7.3.3. *For any square matrices M and N*

$$\operatorname{tr}(MN) = \operatorname{tr}(NM).$$

Example 92 Continuing from the previous example,

$$M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, N = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

so

$$MN = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \neq NM = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

However, $\operatorname{tr}(MN) = 2 + 1 = 3 = 1 + 2 = \operatorname{tr}(NM)$.

Another useful property of the trace is that:

$$\operatorname{tr} M = \operatorname{tr} M^T$$

This is true because the trace only uses the diagonal entries, which are fixed by the transpose. For example,

$$\operatorname{tr} \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} = 4 = \operatorname{tr} \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} = \operatorname{tr} \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}^T.$$

Finally, trace is a linear transformation from matrices to the real numbers. This is easy to check.

7.4 Review Problems

Webwork: | Reading Problems | 2, 3, 4 |

1. Compute the following matrix products

$$\begin{pmatrix} 1 & 2 & 1 \\ 4 & 5 & 2 \\ 7 & 8 & 2 \end{pmatrix} \begin{pmatrix} -2 & \frac{4}{3} & -\frac{1}{3} \\ 2 & -\frac{5}{3} & \frac{2}{3} \\ -1 & 2 & -1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix},$$

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 1 \\ 4 & 5 & 2 \\ 7 & 8 & 2 \end{pmatrix} \begin{pmatrix} -2 & \frac{4}{3} & -\frac{1}{3} \\ 2 & -\frac{5}{3} & \frac{2}{3} \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 4 & 5 & 2 \\ 7 & 8 & 2 \end{pmatrix},$$

$$\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad \begin{pmatrix} 2 & 1 & 2 & 1 & 2 \\ 0 & 2 & 1 & 2 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 2 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 2 & 1 & 2 & 1 \\ 0 & 1 & 2 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\begin{pmatrix} -2 & \frac{4}{3} & -\frac{1}{3} \\ 2 & -\frac{5}{3} & \frac{2}{3} \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 4 & \frac{2}{3} & -\frac{2}{3} \\ 6 & \frac{5}{3} & -\frac{2}{3} \\ 12 & -\frac{16}{3} & \frac{10}{3} \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 4 & 5 & 2 \\ 7 & 8 & 2 \end{pmatrix}.$$

2. Let's prove the theorem $(MN)^T = N^T M^T$.

   Note: the following is a common technique for proving matrix identities.

   (a) Let $M = (m^i_j)$ and let $N = (n^i_j)$. Write out a few of the entries of each matrix in the form given at the beginning of section 7.3.

   (b) Multiply out MN and write out a few of its entries in the same form as in part (a). In terms of the entries of M and the entries of N, what is the entry in row i and column j of MN?

   (c) Take the transpose $(MN)^T$ and write out a few of its entries in the same form as in part (a). In terms of the entries of M and the entries of N, what is the entry in row i and column j of $(MN)^T$?

   (d) Take the transposes N^T and M^T and write out a few of their entries in the same form as in part (a).

   (e) Multiply out $N^T M^T$ and write out a few of its entries in the same form as in part a. In terms of the entries of M and the entries of N, what is the entry in row i and column j of $N^T M^T$?

   (f) Show that the answers you got in parts (c) and (e) are the same.

3. (a) Let $A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 4 \end{pmatrix}$. Find AA^T and $A^T A$ and their traces.

   (b) Let M be any $m \times n$ matrix. Show that $M^T M$ and MM^T are symmetric. (Hint: use the result of the previous problem.) What are their sizes? What is the relationship between their traces?

4. Let $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ be column vectors. Show that the dot product $x \cdot y = x^T \, I \, y$.

Hint

5. Above, we showed that *left* multiplication by an $r \times s$ matrix N was a linear transformation $M^s_k \xrightarrow{N} M^r_k$. Show that *right* multiplication by a $k \times m$ matrix R is a linear transformation $M^s_k \xrightarrow{R} M^s_m$. In other words, show that right matrix multiplication obeys linearity.

Hint

6. Let the V be a vector space where $B = (v_1, v_2)$ is an ordered basis. Suppose

$$L : V \xrightarrow{\text{linear}} V$$

and

$$L(v_1) = v_1 + v_2 \,, \quad L(v_2) = 2v_1 + v_2 \,.$$

Compute the matrix of L in the basis B and then compute the trace of this matrix. Suppose that $ad - bc \neq 0$ and consider now the new basis

$$B' = (av_1 + bv_2, cv_1 + dv_2) \,.$$

Compute the matrix of L in the basis B'. Compute the trace of this matrix. What do you find? What do you conclude about the trace of a matrix? Does it make sense to talk about the "trace of a linear transformation" without reference to any bases?

7. Explain what happens to a matrix when:

   (a) You multiply it on the left by a diagonal matrix.

   (b) You multiply it on the right by a diagonal matrix.

Give a few simple examples before you start explaining.

8. Compute $\exp(A)$ for the following matrices:

   - $A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$
   - $A = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$
   - $A = \begin{pmatrix} 0 & \lambda \\ 0 & 0 \end{pmatrix}$



Hint


9. Let $M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$. Divide M into named blocks, with one block the 4×4 identity matrix, and then multiply blocks to compute M^2.

10. A matrix A is called *anti-symmetric* (or skew-symmetric) if $A^T = -A$. Show that for every $n \times n$ matrix M, we can write $M = A + S$ where A is an anti-symmetric matrix and S is a symmetric matrix.

    Hint: What kind of matrix is $M + M^T$? How about $M - M^T$?

11. An example of an operation which is not associative is the cross product.

    (a) Give a simple example of three vectors from 3-space u, v, w such that $u \times (v \times w) \neq (u \times v) \times w$.

(b) We saw in Chapter 1 that the operator $B = u\times$ (cross product with a vector) is a linear operator. It can therefore be written as a matrix (given an ordered basis such as the standard basis). How is it that composing such linear operators is non-associative even though matrix multiplication is associative?

7.5 Inverse Matrix

Definition A square matrix M is **invertible** (or **nonsingular**) if there exists a matrix M^{-1} such that

$$M^{-1}M = I = MM^{-1}.$$

If M has no inverse, we say M is **singular** or **non-invertible**.

Inverse of a 2×2 Matrix Let M and N be the matrices:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad N = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Multiplying these matrices gives:

$$MN = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} = (ad-bc)I\,.$$

Then $M^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$, so long as $ad-bc \neq 0$.

7.5.1 Three Properties of the Inverse

1. If A is a square matrix and B is the inverse of A, then A is the inverse of B, since $AB = I = BA$. So we have the identity

$$(A^{-1})^{-1} = A.$$

2. Notice that $B^{-1}A^{-1}AB = B^{-1}IB = I = ABB^{-1}A^{-1}$ so

$$\boxed{(AB)^{-1} = B^{-1}A^{-1}}$$

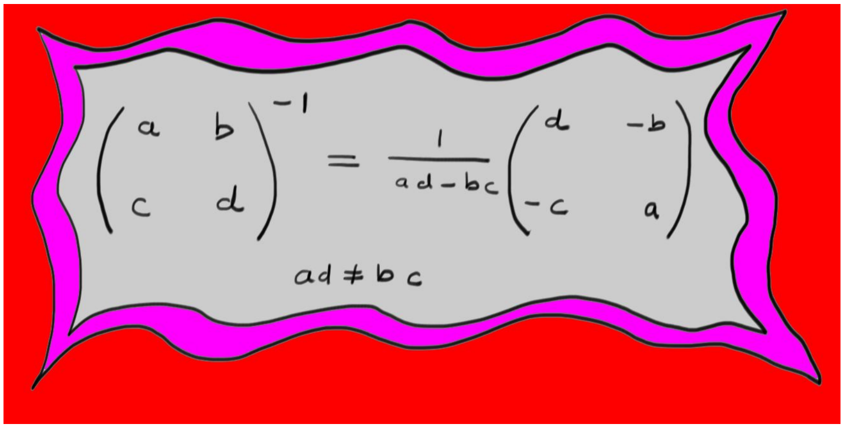


Figure 7.1: The formula for the inverse of a 2×2 matrix is worth memorizing!

Thus, much like the transpose, taking the inverse of a product *reverses* the order of the product.

3. Finally, recall that $(AB)^T = B^T A^T$. Since $I^T = I$, then $(A^{-1}A)^T = A^T(A^{-1})^T = I$. Similarly, $(AA^{-1})^T = (A^{-1})^T A^T = I$. Then:

$$(A^{-1})^T = (A^T)^{-1}$$



2×2 `Example`



7.5.2 Finding Inverses (Redux)

Gaussian elimination can be used to find inverse matrices. This concept is covered in chapter 2, section 2.3.2, but is presented here again as review in more sophisticated terms.

Suppose M is a square invertible matrix and $MX = V$ is a linear system. The solution must be unique because it can be found by multiplying the equation on both sides by M^{-1} yielding $X = M^{-1}V$. Thus, the reduced row echelon form of the linear system has an identity matrix on the left:

$$\big(M \mid V\big) \sim \big(I \mid M^{-1}V\big)$$

Solving the linear system $MX = V$ then tells us what $M^{-1}V$ is.

To solve many linear systems with the same matrix at once,

$$MX = V_1,\ MX = V_2$$

we can consider augmented matrices with many columns on the right and then apply Gaussian row reduction to the left side of the matrix. Once the identity matrix is on the left side of the augmented matrix, then the solution of each of the individual linear systems is on the right.

$$\left(M \mid V_1 \quad V_2 \right) \sim \left(I \mid M^{-1}V_1 \quad M^{-1}V_2 \right)$$

To compute M^{-1}, we would like M^{-1}, rather than $M^{-1}V$ to appear on the right side of our augmented matrix. This is achieved by solving the collection of systems $MX = e_k$, where e_k is the column vector of zeroes with a 1 in the kth entry. *I.e.*, the $n \times n$ identity matrix can be viewed as a bunch of column vectors $I_n = (e_1\ e_2\ \cdots e_n)$. So, putting the e_k's together into an identity matrix, we get:

$$\big(M \mid I\big) \sim \big(I \mid M^{-1}I\big) = \big(I \mid M^{-1}\big)$$

Example 93 Find $\begin{pmatrix} -1 & 2 & -3 \\ 2 & 1 & 0 \\ 4 & -2 & 5 \end{pmatrix}^{-1}$.

We start by writing the augmented matrix, then apply row reduction to the left side.

$$\left(\begin{array}{rrr|rrr} -1 & 2 & -3 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 4 & -2 & 5 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{rrr|rrr} 1 & -2 & 3 & 1 & 0 & 0 \\ 0 & 5 & -6 & 2 & 1 & 0 \\ 0 & 6 & -7 & 4 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{rrr|rrr} 1 & 0 & \frac{3}{5} & -\frac{1}{4} & \frac{2}{5} & 0 \\ 0 & 1 & -\frac{6}{5} & \frac{2}{5} & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} & -\frac{6}{5} & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{rrr|rrr} 1 & 0 & 0 & -5 & 4 & -3 \\ 0 & 1 & 0 & 10 & -7 & 6 \\ 0 & 0 & 1 & 8 & -6 & 5 \end{array}\right)$$

At this point, we know M^{-1} assuming we didn't goof up. However, row reduction is a lengthy and involved process with lots of room for arithmetic errors, so we should *check our answer,* by confirming that $MM^{-1} = I$ (or if you prefer $M^{-1}M = I$):

$$MM^{-1} = \begin{pmatrix} -1 & 2 & -3 \\ 2 & 1 & 0 \\ 4 & -2 & 5 \end{pmatrix} \begin{pmatrix} -5 & 4 & -3 \\ 10 & -7 & 6 \\ 8 & -6 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The product of the two matrices is indeed the identity matrix, so we're done.

Reading homework: problem 5

7.5.3 Linear Systems and Inverses

If M^{-1} exists and is known, then we can immediately solve linear systems associated to M.

Example 94 Consider the linear system:

$$\begin{aligned} -x + 2y - 3z &= 1 \\ 2x + y &= 2 \\ 4x - 2y + 5z &= 0 \end{aligned}$$

The associated matrix equation is $MX = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$, where M is the same as in the previous section, so the system above is equivalent to the matrix equation

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 & 2 & -3 \\ 2 & 1 & 0 \\ 4 & -2 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -5 & 4 & -3 \\ 10 & -7 & 6 \\ 8 & -6 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ -4 \\ -4 \end{pmatrix}.$$

That is, the system is equivalent to the equation $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ -4 \\ -4 \end{pmatrix}$, and it is easy to see what the solution(s) to this equation are.

In summary, when M^{-1} exists

$$Mx = v \Leftrightarrow x = M^{-1}v\,.$$

Reading homework: problem 5

7.5.4 Homogeneous Systems

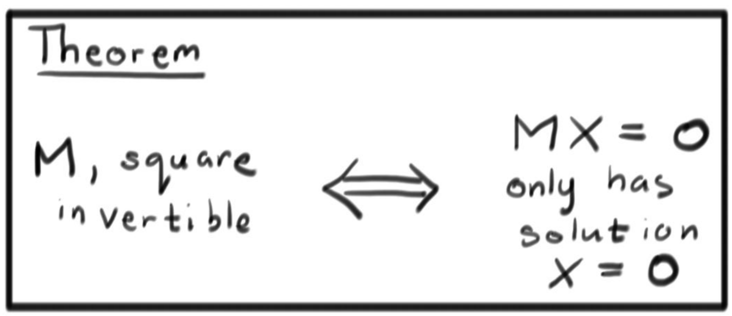
Theorem 7.5.1. *A square matrix M is invertible if and only if the homogeneous system*

$$Mx = 0$$

has no non-zero solutions.

Proof. First, suppose that M^{-1} exists. Then $Mx = 0 \Rightarrow x = M^{-1}0 = 0$. Thus, if M is invertible, then $Mx = 0$ has no non-zero solutions.

On the other hand, $Mx = 0$ always has the solution $x = 0$. If no other solutions exist, then M can be put into reduced row echelon form with every variable a pivot. In this case, M^{-1} can be computed using the process in the previous section. □



7.5.5 Bit Matrices

In computer science, information is recorded using binary strings of data. For example, the following string contains an English word:

011011000110100101101110011001010110000101110010

A *bit* is the basic unit of information, keeping track of a single one or zero. Computers can add and multiply individual bits very quickly.

In chapter 5, section 5.2 it is explained how to formulate vector spaces over fields other than real numbers. In particular, al of the properties of a vector space make sense with numbers $\mathbb{Z}_2 = \{0, 1\}$ with addition and multiplication given by the following tables.

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

Notice that $-1 = 1$, since $1+1 = 0$. Therefore, we can apply all of the linear algebra we have learned thus far to matrices with $\mathbb{Z}_2$ entries. A matrix with entries in $\mathbb{Z}_2$ is sometimes called a *bit matrix*.

Example 95 $\begin{pmatrix}1&0&1\\0&1&1\\1&1&1\end{pmatrix}$ is an invertible matrix over $\mathbb{Z}_2$;

$$\begin{pmatrix}1&0&1\\0&1&1\\1&1&1\end{pmatrix}^{-1} = \begin{pmatrix}0&1&1\\1&0&1\\1&1&1\end{pmatrix}.$$

This can be easily verified by multiplying:

$$\begin{pmatrix}1&0&1\\0&1&1\\1&1&1\end{pmatrix}\begin{pmatrix}0&1&1\\1&0&1\\1&1&1\end{pmatrix} = \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$$

Application: Cryptography A very simple way to hide information is to use a substitution cipher, in which the alphabet is permuted and each letter in a message is systematically exchanged for another. For example, the ROT-13 cypher just exchanges a letter with the letter thirteen places before or after it in the alphabet. For example, HELLO becomes URYYB. Applying the algorithm again decodes the message, turning URYYB back into HELLO. Substitution ciphers are easy to break, but the basic idea can be extended to create cryptographic systems that are practically uncrackable. For example, a *one-time pad* is a system that uses a different substitution for each letter in the message. So long as a particular set of substitutions is not used on more than one message, the one-time pad is unbreakable.

English characters are often stored in computers in the ASCII format. In ASCII, a single character is represented by a string of eight bits, which we can consider as a vector in $\mathbb{Z}_2^8$ (which is like vectors in $\mathbb{R}^8$, where the entries are zeros and ones). One way to create a substitution cipher, then, is to choose an 8×8 invertible bit matrix M, and multiply each letter of the message by M. Then to decode the message, each string of eight characters would be multiplied by M^{-1}.

To make the message a bit tougher to decode, one could consider pairs (or longer sequences) of letters as a single vector in $\mathbb{Z}_2^{16}$ (or a higher-dimensional space), and then use an appropriately-sized invertible matrix. For more on cryptography, see "The Code Book," by Simon Singh (1999, Doubleday).

7.6 Review Problems

Webwork:

Reading Problems	6, 7

1. Find formulas for the inverses of the following matrices, when they are not singular:

   (a) $\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}$

   (b) $\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$

   When are these matrices singular?

2. Write down all 2×2 bit matrices and decide which of them are singular. For those which are not singular, pair them with their inverse.

3. Let M be a square matrix. Explain why the following statements are equivalent:

   (a) $MX = V$ has a *unique* solution for every column vector V.

   (b) M is non-singular.

   Hint: In general for problems like this, think about the key words:

   First, suppose that there is some column vector V such that the equation $MX = V$ has two distinct solutions. Show that M must be singular; that is, show that M can have no inverse.

   Next, suppose that there is some column vector V such that the equation $MX = V$ has no solutions. Show that M must be singular.

   Finally, suppose that M is non-singular. Show that no matter what the column vector V is, there is a unique solution to $MX = V$.



Hint



4. *Left and Right Inverses:* So far we have only talked about inverses of square matrices. This problem will explore the notion of a left and right inverse for a matrix that is not square. Let

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

(a) Compute:

i. AA^T,

ii. $\left(AA^T\right)^{-1}$,

iii. $B := A^T\left(AA^T\right)^{-1}$

(b) Show that the matrix B above is a *right inverse* for A, *i.e.*, verify that

$$AB = I\,.$$

(c) Is BA defined? (Why or why not?)

(d) Let A be an $n \times m$ matrix with $n > m$. Suggest a formula for a left inverse C such that

$$CA = I$$

Hint: you may assume that $A^T A$ has an inverse.

(e) Test your proposal for a left inverse for the simple example

$$A = \begin{pmatrix} 1 \\ 2 \end{pmatrix}\,,$$

(f) True or false: Left and right inverses are unique. If false give a counterexample.

 Hint 

5. Show that if the range (remember that the range of a function is the set of all its outputs, not the codomain) of a 3×3 matrix M (viewed as a function $\mathbb{R}^3 \to \mathbb{R}^3$) is a plane then one of the columns is a sum of multiples of the other columns. Show that this relationship is preserved under EROs. Show, further, that the solutions to $Mx = 0$ describe this relationship between the columns.

6. If M and N are square matrices of the same size such that M^{-1} exists and N^{-1} does not exist, does $(MN)^{-1}$ exist?

7. If M is a square matrix which is not invertible, is e^M invertible?

8. Elementary Column Operations (ECOs) can be defined in the same 3 types as EROs. Describe the 3 kinds of ECOs. Show that if maximal elimination using ECOs is performed on a square matrix and a column of zeros is obtained then that matrix is not invertible.

7.7 LU Redux

Certain matrices are easier to work with than others. In this section, we will see how to write any square[2] matrix M as the product of two simpler matrices. We will write

$$M = LU\,,$$

where:

- L is *lower triangular*. This means that all entries above the main diagonal are zero. In notation, $L = (l^i_j)$ with $l^i_j = 0$ for all $j > i$.

$$L = \begin{pmatrix} l^1_1 & 0 & 0 & \cdots \\ l^2_1 & l^2_2 & 0 & \cdots \\ l^3_1 & l^3_2 & l^3_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

- U is *upper triangular*. This means that all entries below the main diagonal are zero. In notation, $U = (u^i_j)$ with $u^i_j = 0$ for all $j < i$.

$$U = \begin{pmatrix} u^1_1 & u^1_2 & u^1_3 & \cdots \\ 0 & u^2_2 & u^2_3 & \cdots \\ 0 & 0 & u^3_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$M = LU$ is called an LU *decomposition* of M.

This is a useful trick for computational reasons; it is much easier to compute the inverse of an upper or lower triangular matrix than general matrices. Since inverses are useful for solving linear systems, this makes solving any linear system associated to the matrix much faster as well. The determinant—a very important quantity associated with any square matrix—is very easy to compute for triangular matrices.

Example 96 Linear systems associated to upper triangular matrices are very easy to solve by back substitution.

$$\left(\begin{array}{cc|c} a & b & 1 \\ 0 & c & e \end{array}\right) \Rightarrow y = \frac{e}{c}\,, \quad x = \frac{1}{a}\left(1 - \frac{be}{c}\right)$$

[2]The case where M is not square is dealt with at the end of the section.

$$\begin{pmatrix} 1 & 0 & 0 & d \\ a & 1 & 0 & e \\ b & c & 1 & f \end{pmatrix} \Rightarrow \left\{ \begin{array}{l} x = d \\ y = e - ax \\ z = f - bx - cy \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} x = d \\ y = e - ad \\ z = f - bd - c(e - ad) \end{array} \right. .$$

For lower triangular matrices, *forward* substitution gives a quick solution; for upper triangular matrices, *back* substitution gives the solution.

7.7.1 Using LU Decomposition to Solve Linear Systems

Suppose we have $M = LU$ and want to solve the system

$$MX = LUX = V.$$

- Step 1: Set $W = \begin{pmatrix} u \\ v \\ w \end{pmatrix} = UX$.
- Step 2: Solve the system $LW = V$. This should be simple by forward substitution since L is lower triangular. Suppose the solution to $LW = V$ is W_0.
- Step 3: Now solve the system $UX = W_0$. This should be easy by backward substitution, since U is upper triangular. The solution to this system is the solution to the original system.

We can think of this as using the matrix L to perform row operations on the matrix U in order to solve the system; this idea also appears in the study of determinants.

Reading homework: problem 7

Example 97 Consider the linear system:

$$\begin{aligned} 6x + 18y + 3z &= 3 \\ 2x + 12y + z &= 19 \\ 4x + 15y + 3z &= 0 \end{aligned}$$

An LU decomposition for the associated matrix M is

$$\begin{pmatrix} 6 & 18 & 3 \\ 2 & 12 & 1 \\ 4 & 15 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 6 & 0 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 6 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- Step 1: Set $W = \begin{pmatrix} u \\ v \\ w \end{pmatrix} = UX$.

- Step 2: Solve the system $LW = V$:

$$\begin{pmatrix} 3 & 0 & 0 \\ 1 & 6 & 0 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 3 \\ 19 \\ 0 \end{pmatrix}$$

By substitution, we get $u = 1$, $v = 3$, and $w = -11$. Then

$$W_0 = \begin{pmatrix} 1 \\ 3 \\ -11 \end{pmatrix}$$

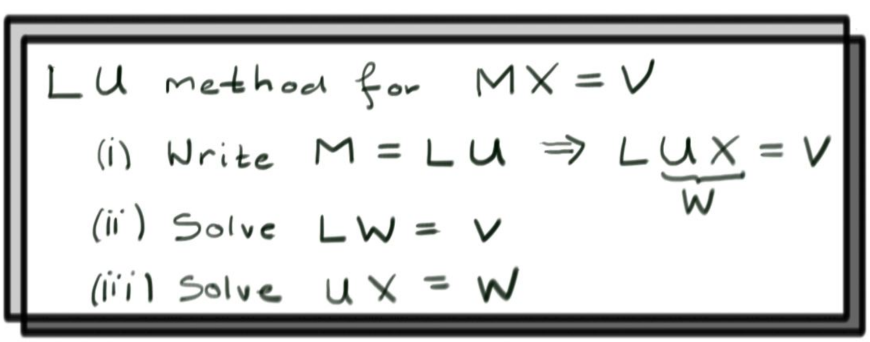
- Step 3: Solve the system $UX = W_0$.

$$\begin{pmatrix} 2 & 6 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ -11 \end{pmatrix}$$

Back substitution gives $z = -11, y = 3$, and $x = -3$.

Then $X = \begin{pmatrix} -3 \\ 3 \\ -11 \end{pmatrix}$, and we're done.

Using an LU decomposition



7.7.2 Finding an LU Decomposition.

In chapter 2, section 2.3.4, Gaussian elimination was used to find LU matrix decompositions. These ideas are presented here again as review.

For any given matrix, there are actually many different LU decompositions. However, there is a unique LU decomposition in which the L matrix has ones on the diagonal. In that case L is called a *lower unit triangular matrix*.

To find the LU decomposition, we'll create two sequences of matrices $L_1, L_2, \ldots$ and $U_1, U_2, \ldots$ such that at each step, $L_iU_i = M$. Each of the L_i will be lower triangular, but only the last U_i will be upper triangular. The main trick for this calculation is captured by the following example:

Example 98 (An Elementary Matrix)
Consider

$$E = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}, \qquad M = \begin{pmatrix} a & b & c & \cdots \\ d & e & f & \cdots \end{pmatrix}.$$

Lets compute EM

$$EM = \begin{pmatrix} a & b & c & \cdots \\ d+\lambda a & e+\lambda b & f+\lambda c & \cdots \end{pmatrix}.$$

Something neat happened here: multiplying M by E performed the row operation $R_2 \to R_2 + \lambda R_1$ on M. Another interesting fact:

$$E^{-1} := \begin{pmatrix} 1 & 0 \\ -\lambda & 1 \end{pmatrix}$$

obeys (check this yourself...)

$$E^{-1}E = 1.$$

Hence $M = E^{-1}EM$ or, writing this out

$$\begin{pmatrix} a & b & c & \cdots \\ d & e & f & \cdots \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\lambda & 1 \end{pmatrix} \begin{pmatrix} a & b & c & \cdots \\ d+\lambda a & e+\lambda b & f+\lambda c & \cdots \end{pmatrix}.$$

Here the matrix on the left is lower triangular, while the matrix on the right has had a row operation performed on it.

We would like to use the first row of M to zero out the first entry of every row below it. For our running example,

$$M = \begin{pmatrix} 6 & 18 & 3 \\ 2 & 12 & 1 \\ 4 & 15 & 3 \end{pmatrix},$$

so we would like to perform the row operations

$$R_2 \to R_2 - \frac{1}{3}R_1 \text{ and } R_3 \to R_3 - \frac{2}{3}R_1 \,.$$

If we perform these row operations on M to produce

$$U_1 = \begin{pmatrix} 6 & 18 & 3 \\ 0 & 6 & 0 \\ 0 & 3 & 1 \end{pmatrix} ,$$

we need to multiply this on the left by a lower triangular matrix L_1 so that the product $L_1U_1 = M$ still. The above example shows how to do this: Set L_1 to be the lower triangular matrix whose first column is filled with **minus** the constants used to zero out the first column of M. Then

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & 0 & 1 \end{pmatrix} .$$

By construction $L_1U_1 = M$, but you should compute this yourself as a double check.

Now repeat the process by zeroing the second column of U_1 below the diagonal using the second row of U_1 using the row operation $R_3 \to R_3 - \frac{1}{2}R_2$ to produce

$$U_2 = \begin{pmatrix} 6 & 18 & 3 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

The matrix that undoes this row operation is obtained in the same way we found L_1 above and is:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & 1 \end{pmatrix} .$$

Thus our answer for L_2 is the product of this matrix with L_1, namely

$$L_2 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & \frac{1}{2} & 1 \end{pmatrix} .$$

Notice that it is lower triangular because

> The product of lower triangular matrices is always lower triangular!

Moreover it is obtained by recording minus the constants used for all our row operations in the appropriate columns (this always works this way). Moreover, U_2 is upper triangular and $M = L_2 U_2$, we are done! Putting this all together we have

$$M = \begin{pmatrix} 6 & 18 & 3 \\ 2 & 12 & 1 \\ 4 & 15 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 6 & 18 & 3 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

If the matrix you're working with has more than three rows, just continue this process by zeroing out the next column below the diagonal, and repeat until there's nothing left to do.

Another LU decomposition example

The fractions in the L matrix are admittedly ugly. For two matrices LU, we can multiply one entire column of L by a constant λ and divide the corresponding row of U by the same constant without changing the product of the two matrices. Then:

$$\begin{aligned} LU &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & \frac{1}{2} & 1 \end{pmatrix} I \begin{pmatrix} 6 & 18 & 3 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{2}{3} & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{6} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 6 & 18 & 3 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 3 & 0 & 0 \\ 1 & 6 & 0 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 6 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

The resulting matrix looks nicer, but isn't in standard (lower unit triangular matrix) form.

Reading homework: problem 7

For matrices that are not square, LU decomposition still makes sense. Given an $m \times n$ matrix M, for example we could write $M = LU$ with L a square lower unit triangular matrix, and U a rectangular matrix. Then L will be an $m \times m$ matrix, and U will be an $m \times n$ matrix (of the same shape as M). From here, the process is exactly the same as for a square matrix. We create a sequence of matrices L_i and U_i that is eventually the LU decomposition. Again, we start with $L_0 = I$ and $U_0 = M$.

Example 99 Let's find the LU decomposition of $M = U_0 = \begin{pmatrix} -2 & 1 & 3 \\ -4 & 4 & 1 \end{pmatrix}$. Since M is a 2×3 matrix, our decomposition will consist of a 2×2 matrix and a 2×3 matrix. Then we start with $L_0 = I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

The next step is to zero-out the first column of M below the diagonal. There is only one row to cancel, then, and it can be removed by subtracting 2 times the first row of M to the second row of M. Then:

$$L_1 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \qquad U_1 = \begin{pmatrix} -2 & 1 & 3 \\ 0 & 2 & -5 \end{pmatrix}$$

Since U_1 is upper triangular, we're done. With a larger matrix, we would just continue the process.

7.7.3 Block LDU Decomposition

Let M be a square block matrix with square blocks X, Y, Z, W such that X^{-1} exists. Then M can be decomposed as a block LDU decomposition, where D is block diagonal, as follows:

$$M = \begin{pmatrix} X & Y \\ Z & W \end{pmatrix}$$

Then:

$$M = \begin{pmatrix} I & 0 \\ ZX^{-1} & I \end{pmatrix} \begin{pmatrix} X & 0 \\ 0 & W - ZX^{-1}Y \end{pmatrix} \begin{pmatrix} I & X^{-1}Y \\ 0 & I \end{pmatrix}.$$

This can be checked explicitly simply by block-multiplying these three matrices.



Block LDU Explanation

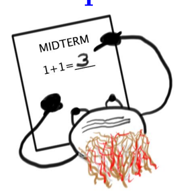


Example 100 For a 2×2 matrix, we can regard each entry as a 1×1 block.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

By multiplying the diagonal matrix by the upper triangular matrix, we get the standard LU decomposition of the matrix.

You are now ready to attempt the first sample midterm.



7.8 Review Problems

Webwork:

Reading Problems	7, 8
LU Decomposition	14

1. Consider the linear system:

$$\begin{array}{llll} x^1 & & & = v^1 \\ l_1^2 x^1 & +x^2 & & = v^2 \\ \vdots & & & \vdots \\ l_1^n x^1 & +l_2^n x^2 & + \cdots + x^n & = v^n \end{array}$$

(i) Find x^1.

(ii) Find x^2.

(iii) Find x^3.

(k) Try to find a formula or recursive method for finding x^k. Don't worry about simplifying your answer.

2. Let $M = \begin{pmatrix} X & Y \\ Z & W \end{pmatrix}$ be a square $n \times n$ block matrix with W invertible.

   i. If W has r rows, what size are X, Y, and Z?

   ii. Find a UDL decomposition for M. In other words, fill in the stars in the following equation:

$$\begin{pmatrix} X & Y \\ Z & W \end{pmatrix} = \begin{pmatrix} I & * \\ 0 & I \end{pmatrix} \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \begin{pmatrix} I & 0 \\ * & I \end{pmatrix}$$

3. Show that if M is a square matrix which is not invertible then either the matrix matrix U or the matrix L in the LU-decomposition $M = LU$ has a zero on it's diagonal.

4. Describe what upper and lower triangular matrices do to the unit hypercube in their domain.

5. In chapter 3 we saw that, since in general row exchange matrices are necessary to achieve upper triangular form, $LDPU$ factorization is the complete decomposition of an invertible matrix into EROs of various kinds. Suggest a procedure for using $LDPU$ decompositions to solve linear systems that generalizes the procedure above.

6. Is there a reason to prefer LU decomposition to UL decomposition, or is the order just a convention?

7. If M is invertible then what are the LU, LDU, and $LDPU$ decompositions of M^T in terms of the decompositions for M? Can you do the same for M^{-1}?

8. Argue that if M is symmetric then $L = U^T$ in the LDU decomposition of M.

8 Determinants

Given a square matrix, is there an easy way to know when it is invertible? Answering this fundamental question is the goal of this chapter.

8.1 The Determinant Formula

The determinant boils down a square matrix to a a single number. That number determines whether the square matrix is invertible or not. Lets see how this works for small matrices first.

8.1.1 Simple Examples

For small cases, we already know when a matrix is invertible. If M is a 1×1 matrix, then $M = (m) \Rightarrow M^{-1} = (1/m)$. Then M is invertible if and only if $m \neq 0$.

For M a 2×2 matrix, chapter 7 section 7.5 shows that if

$$M = \begin{pmatrix} m_1^1 & m_2^1 \\ m_1^2 & m_2^2 \end{pmatrix},$$

then

$$M^{-1} = \frac{1}{m_1^1 m_2^2 - m_2^1 m_1^2} \begin{pmatrix} m_2^2 & -m_2^1 \\ -m_1^2 & m_1^1 \end{pmatrix}.$$

Thus M is invertible if and only if

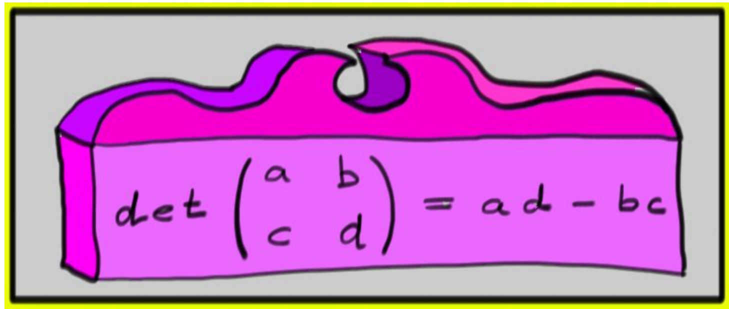


Figure 8.1: Memorize the determinant formula for a 2×2 matrix!

$$m^1_1 m^2_2 - m^1_2 m^2_1 \neq 0\,.$$

For 2×2 matrices, this quantity is called the *determinant of* M.

$$\det M = \det \begin{pmatrix} m^1_1 & m^1_2 \\ m^2_1 & m^2_2 \end{pmatrix} = m^1_1 m^2_2 - m^1_2 m^2_1\,.$$

Example 101 For a 3×3 matrix,

$$M = \begin{pmatrix} m^1_1 & m^1_2 & m^1_3 \\ m^2_1 & m^2_2 & m^2_3 \\ m^3_1 & m^3_2 & m^3_3 \end{pmatrix}\,,$$

then—see review question 1—M is non-singular if and only if:

$$\det M = m^1_1 m^2_2 m^3_3 - m^1_1 m^2_3 m^3_2 + m^1_2 m^2_3 m^3_1 - m^1_2 m^2_1 m^3_3 + m^1_3 m^2_1 m^3_2 - m^1_3 m^2_2 m^3_1 \neq 0.$$

Notice that in the subscripts, each ordering of the numbers 1, 2, and 3 occurs exactly once. Each of these is a *permutation* of the set $\{1, 2, 3\}$.

8.1.2 Permutations

Consider n objects labeled 1 through n and shuffle them. Each possible shuffle is called a *permutation*. For example, here is an example of a permutation of 1–5:

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 5 & 1 & 3 \end{bmatrix}$$

We can consider a permutation σ as an invertible function from the set of numbers $[n] := \{1, 2, \dots, n\}$ to $[n]$, so can write $\sigma(3) = 5$ in the above example. In general we can write

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ \sigma(1) & \sigma(2) & \sigma(3) & \sigma(4) & \sigma(5) \end{bmatrix},$$

but since the top line of any permutation is always the same, we can omit it and just write:

$$\sigma = \begin{bmatrix} \sigma(1) & \sigma(2) & \sigma(3) & \sigma(4) & \sigma(5) \end{bmatrix}$$

and so our example becomes simply $\sigma = [4\,2\,5\,1\,3]$.

The mathematics of permutations is extensive; there are a few key properties of permutations that we'll need:

- There are $n!$ permutations of n distinct objects, since there are n choices for the first object, $n - 1$ choices for the second once the first has been chosen, and so on.

- Every permutation can be built up by successively swapping pairs of objects. For example, to build up the permutation $\begin{bmatrix} 3 & 1 & 2 \end{bmatrix}$ from the trivial permutation $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, you can first swap 2 and 3, and then swap 1 and 3.

- For any given permutation σ, there is some number of swaps it takes to build up the permutation. (It's simplest to use the minimum number of swaps, but you don't have to: it turns out that *any* way of building up the permutation from swaps will have have the same parity of swaps, either even or odd.) If this number happens to be even, then σ is called an *even permutation*; if this number is odd, then σ is an *odd permutation*. In fact, $n!$ is even for all $n \geq 2$, and exactly half of the permutations are even and the other half are odd. It's worth noting that the trivial permutation (which sends $i \to i$ for every i) is an even permutation, since it uses zero swaps.

Definition The **sign function** is a function sgn that sends permutations to the set $\{-1, 1\}$ with rule of correspondence defined by

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd.} \end{cases}$$

Permutation Example

Reading homework: problem 1

We can use permutations to give a definition of the determinant.

Definition The **determinant** of $n \times n$ matrix M is

$$\det M = \sum_{\sigma} \operatorname{sgn}(\sigma)\, m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)}.$$

The sum is over all permutations of n objects; a sum over the all elements of $\{\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}\}$. Each summand is a product of n entries from the matrix with each factor from a different row. In different terms of the sum the column numbers are shuffled by different permutations σ.

The last statement about the summands yields a nice property of the determinant:

Theorem 8.1.1. *If $M = (m^i_j)$ has a row consisting entirely of zeros, then $m^i_{\sigma(i)} = 0$ for every σ and some i. Moreover $\det M = 0$.*

Example 102 Because there are many permutations of n, writing the determinant this way for a general matrix gives a very long sum. For $n = 4$, there are $24 = 4!$ permutations, and for $n = 5$, there are already $120 = 5!$ permutations.

For a 4×4 matrix, $M = \begin{pmatrix} m^1_1 & m^1_2 & m^1_3 & m^1_4 \\ m^2_1 & m^2_2 & m^2_3 & m^2_4 \\ m^3_1 & m^3_2 & m^3_3 & m^3_4 \\ m^4_1 & m^4_2 & m^4_3 & m^4_4 \end{pmatrix}$, then $\det M$ is:

$$\begin{aligned} \det M &= m^1_1 m^2_2 m^3_3 m^4_4 - m^1_1 m^2_3 m^3_2 m^4_4 - m^1_1 m^2_2 m^3_4 m^4_3 \\ &\quad - m^1_2 m^2_1 m^3_3 m^4_4 + m^1_1 m^2_3 m^3_4 m^4_2 + m^1_1 m^2_4 m^3_2 m^4_3 \\ &\quad + m^1_2 m^2_3 m^3_1 m^4_4 + m^1_2 m^2_1 m^3_4 m^4_3 \pm \text{16 more terms.} \end{aligned}$$

This is very cumbersome.

Luckily, it is very easy to compute the determinants of certain matrices. For example, if M is diagonal, meaning that $M^i_j = 0$ whenever $i \neq j$, then all summands of the determinant involving off-diagonal entries vanish and

$$\det M = \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)} = m^1_1 m^2_2 \cdots m^n_n .$$

The determinant of a diagonal matrix is the product of its diagonal entries.

Since the identity matrix is diagonal with all diagonal entries equal to one, we have

$$\det I = 1 .$$

We would like to use the determinant to decide whether a matrix is invertible. Previously, we computed the inverse of a matrix by applying row operations. Therefore we ask what happens to the determinant when row operations are applied to a matrix.

Swapping rows Lets swap rows i and j of a matrix M and then compute its determinant. For the permutation σ, let $\hat{\sigma}$ be the permutation obtained by swapping positions i and j. Clearly

$$\operatorname{sgn}(\hat{\sigma}) = -\operatorname{sgn}(\sigma) .$$

Let M' be the matrix M with rows i and j swapped. Then (assuming $i < j$):

$$\begin{aligned}
\det M' &= \sum_\sigma \operatorname{sgn}(\sigma)\, m^1_{\sigma(1)} \cdots m^j_{\sigma(i)} \cdots m^i_{\sigma(j)} \cdots m^n_{\sigma(n)} \\
&= \sum_\sigma \operatorname{sgn}(\sigma)\, m^1_{\sigma(1)} \cdots m^i_{\sigma(j)} \cdots m^j_{\sigma(i)} \cdots m^n_{\sigma(n)} \\
&= \sum_\sigma (-\operatorname{sgn}(\hat{\sigma}))\, m^1_{\hat{\sigma}(1)} \cdots m^i_{\hat{\sigma}(i)} \cdots m^j_{\hat{\sigma}(j)} \cdots m^n_{\hat{\sigma}(n)} \\
&= -\sum_{\hat{\sigma}} \operatorname{sgn}(\hat{\sigma})\, m^1_{\hat{\sigma}(1)} \cdots m^i_{\hat{\sigma}(i)} \cdots m^j_{\hat{\sigma}(j)} \cdots m^n_{\hat{\sigma}(n)} \\
&= -\det M .
\end{aligned}$$

The step replacing $\sum_\sigma$ by $\sum_{\hat{\sigma}}$ often causes confusion; it holds since we sum over all permutations (see review problem 3). Thus we see that swapping rows changes the sign of the determinant. *I.e.*,

$$\det M' = -\det M .$$

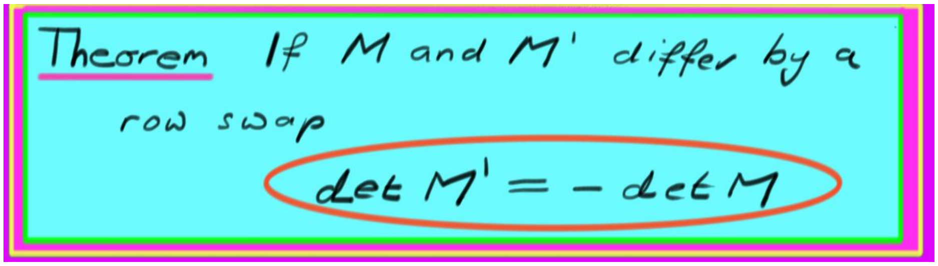


Figure 8.2: Remember what row swap does to determinants!

Reading homework: problem 8.2

Applying this result to $M = I$ (the identity matrix) yields

$$\det E^i_j = -1\,,$$

where the matrix E^i_j is the identity matrix with rows i and j swapped. It is a row swap elementary matrix.

This implies another nice property of the determinant. If two rows of the matrix are identical, then swapping the rows changes the sign of the matrix, but leaves the matrix unchanged. Then we see the following:

Theorem 8.1.2. *If M has two identical rows, then* $\det M = 0$.

8.2 Elementary Matrices and Determinants

In chapter 2 we found the matrices that perform the row operations involved in Gaussian elimination; we called them elementary matrices.

As a reminder, for any matrix M, and a matrix M' equal to M after a row operation, multiplying by an elementary matrix E gave $M' = EM$.



`Elementary Matrices`



We now examine what the elementary matrices to do determinants.

8.2.1 Row Swap

Our first elementary matrix swaps rows i and j when it is applied to a matrix M. Explicitly, let R^1 through R^n denote the rows of M, and let M' be the matrix M with rows i and j swapped. Then M and M' can be regarded as a block matrices (where the blocks are rows);

$$M = \begin{pmatrix} \vdots \\ R^i \\ \vdots \\ R^j \\ \vdots \end{pmatrix} \text{ and } M' = \begin{pmatrix} \vdots \\ R^j \\ \vdots \\ R^i \\ \vdots \end{pmatrix}.$$

Then notice that

$$M' = \begin{pmatrix} \vdots \\ R^j \\ \vdots \\ R^i \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{pmatrix} \vdots \\ R^i \\ \vdots \\ R^j \\ \vdots \end{pmatrix}.$$

The matrix

$$\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} =: E^i_j$$

is just the identity matrix with rows i and j swapped. The matrix E^i_j is an *elementary matrix* and

$$M' = E^i_j M\,.$$

Because $\det I = 1$ and swapping a pair of rows changes the sign of the determinant, we have found that

$$\det E^i_j = -1\,.$$

Now we know that swapping a pair of rows flips the sign of the determinant so $\det M' = -detM$. But $\det E^i_j = -1$ and $M' = E^i_j M$ so

$$\det E^i_j M = \det E^i_j \, \det M\,.$$

This result hints at a general rule for determinants of products of matrices.

8.2.2 Row Multiplication

The next row operation is multiplying a row by a scalar. Consider

$$M = \begin{pmatrix} R^1 \\ \vdots \\ R^n \end{pmatrix},$$

where R^i are row vectors. Let $R^i(\lambda)$ be the identity matrix, with the ith diagonal entry replaced by λ, not to be confused with the row vectors. *I.e.*,

$$R^i(\lambda) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}.$$

Then:

$$M' = R^i(\lambda)M = \begin{pmatrix} R^1 \\ \vdots \\ \lambda R^i \\ \vdots \\ R^n \end{pmatrix},$$

equals M with one row multiplied by λ.

What effect does multiplication by the elementary matrix $R^i(\lambda)$ have on the determinant?

$$\begin{aligned}
\det M' &= \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots \lambda m^i_{\sigma(i)} \cdots m^n_{\sigma(n)} \\
&= \lambda \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots m^i_{\sigma(i)} \cdots m^n_{\sigma(n)} \\
&= \lambda \det M
\end{aligned}$$

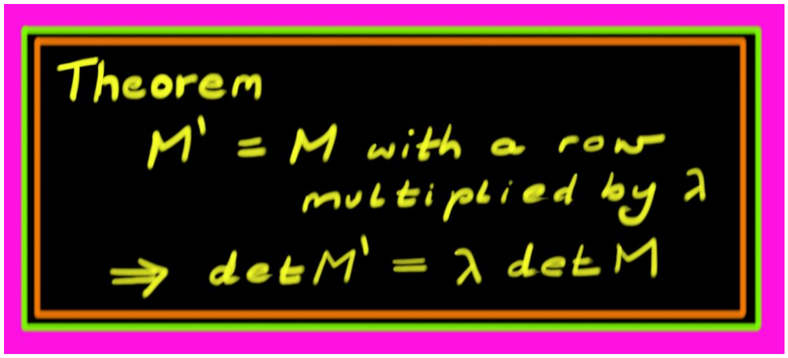


Figure 8.3: Rescaling a row rescales the determinant.

Thus, multiplying a row by λ multiplies the determinant by λ. *I.e.,*

$$\det R^i(\lambda)M = \lambda \det M\,.$$

Since $R^i(\lambda)$ is just the identity matrix with a single row multiplied by λ, then by the above rule, the determinant of $R^i(\lambda)$ is λ. Thus

$$\det R^i(\lambda) = \det \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} = \lambda\,,$$

and once again we have a product of determinants formula

$$\det\big(R^i(\lambda)M\big) = \det\big(R^i(\lambda)\big)\det M.$$

8.2.3 Row Addition

The final row operation is adding μR^j to R^i. This is done with the elementary matrix $S^i_j(\mu)$, which is an identity matrix but with an additional μ in the i,j position;

$$S^i_j(\mu) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & \mu & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}.$$

Then multiplying M by $S^i_j(\mu)$ performs a row addition;

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & \mu & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{pmatrix} \vdots \\ R^i \\ \vdots \\ R^j \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ R^i + \mu R^j \\ \vdots \\ R^j \\ \vdots \end{pmatrix}.$$

What is the effect of multiplying by $S^i_j(\mu)$ on the determinant? Let $M' = S^i_j(\mu)M$, and let M'' be the matrix M but with R^i replaced by R^j Then

$$\begin{aligned} \det M' &= \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots (m^i_{\sigma(i)} + \mu m^j_{\sigma(i)}) \cdots m^n_{\sigma(n)} \\ &= \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots m^i_{\sigma(i)} \cdots m^n_{\sigma(n)} \\ &\qquad + \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} \cdots \mu m^j_{\sigma(j)} \cdots m^j_{\sigma(j)} \cdots m^n_{\sigma(n)} \\ &= \det M + \mu \det M'' \end{aligned}$$

Since M'' has two identical rows, its determinant is 0 so

$$\det M' = \det M,$$

when M' is obtained from M by adding μ times row j to row i.

Reading homework: problem 3

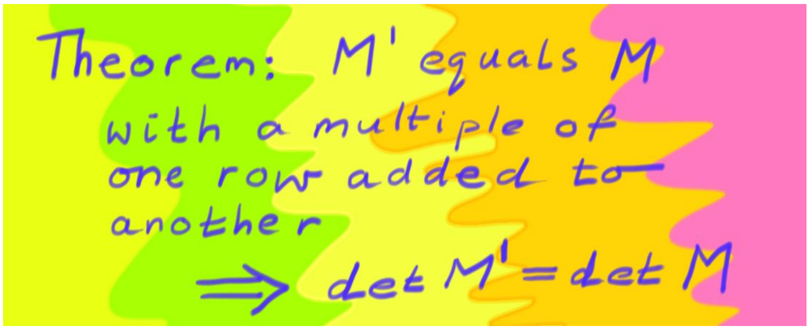


Figure 8.4: Adding one row to another leaves the determinant unchanged.

We also have learnt that

$$\det\left(S^i_j(\mu)M\right) = \det M\,.$$

Notice that if M is the identity matrix, then we have

$$\det S^i_j(\mu) = \det(S^i_j(\mu)I) = \det I = 1\,.$$

8.2.4 Determinant of Products

In summary, the elementary matrices for each of the row operations obey

$$\begin{array}{lcll} E^i_j & = & I \text{ with rows } i,j \text{ swapped;} & \det E^i_j = -1 \\ R^i(\lambda) & = & I \text{ with } \lambda \text{ in position } i,i; & \det R^i(\lambda) = \lambda \\ S^i_j(\mu) & = & I \text{ with } \mu \text{ in position } i,j; & \det S^i_j(\mu) = 1 \end{array}$$



`Elementary Determinants`


Moreover we found a useful formula for determinants of products:

Theorem 8.2.1. *If E is* any *of the elementary matrices $E^i_j, R^i(\lambda), S^i_j(\mu)$, then* $\det(EM) = \det E \det M$*.*

We have seen that any matrix M can be put into reduced row echelon form via a sequence of row operations, and we have seen that any row operation can be achieved via left matrix multiplication by an elementary matrix. Suppose that $\mathrm{RREF}(M)$ is the reduced row echelon form of M. Then

$$\mathrm{RREF}(M) = E_1 E_2 \cdots E_k M\,,$$

where each E_i is an elementary matrix. We know how to compute determinants of elementary matrices and products thereof, so we ask:

What is the determinant of a square matrix in reduced row echelon form?

The answer has two cases:

1. If M is not invertible, then some row of $\mathrm{RREF}(M)$ contains only zeros. Then we can multiply the zero row by any constant λ without changing M; by our previous observation, this scales the determinant of M by λ. Thus, if M is not invertible, $\det \mathrm{RREF}(M) = \lambda \det \mathrm{RREF}(M)$, and so $\det \mathrm{RREF}(M) = 0$.

2. Otherwise, every row of $\mathrm{RREF}(M)$ has a pivot on the diagonal; since M is square, this means that $\mathrm{RREF}(M)$ is the identity matrix. So if M is invertible, $\det \mathrm{RREF}(M) = 1$.

Notice that because $\det \mathrm{RREF}(M) = \det(E_1 E_2 \cdots E_k M)$, by the theorem above,

$$\det \mathrm{RREF}(M) = \det(E_1) \cdots \det(E_k) \det M\,.$$

Since each E_i has non-zero determinant, then $\det \mathrm{RREF}(M) = 0$ if and only if $\det M = 0$. This establishes an important theorem:

Theorem 8.2.2. *For any square matrix M, $\det M \neq 0$ if and only if M is invertible.*

Since we know the determinants of the elementary matrices, we can immediately obtain the following:

`Determinants and Inverses`



Figure 8.5: Determinants measure if a matrix is invertible.

Corollary 8.2.3. *Any elementary matrix $E^i_j, R^i(\lambda), S^i_j(\mu)$ is invertible, except for $R^i(0)$. In fact, the inverse of an elementary matrix is another elementary matrix.*

To obtain one last important result, suppose that M and N are square $n \times n$ matrices, with reduced row echelon forms such that, for elementary matrices E_i and F_i,

$$M = E_1 E_2 \cdots E_k \text{ RREF}(M)\,,$$

and

$$N = F_1 F_2 \cdots F_l \text{ RREF}(N)\,.$$

If RREF(M) is the identity matrix (*i.e.*, M is invertible), then:

$$\begin{aligned}
\det(MN) &= \det(E_1 E_2 \cdots E_k \text{ RREF}(M) F_1 F_2 \cdots F_l \text{ RREF}(N)) \\
&= \det(E_1 E_2 \cdots E_k I F_1 F_2 \cdots F_l \text{ RREF}(N)) \\
&= \det(E_1) \cdots \det(E_k) \det(I) \det(F_1) \cdots \det(F_l) \det \text{RREF}(N) \\
&= \det(M) \det(N)
\end{aligned}$$

Otherwise, M is not invertible, and $\det M = 0 = \det \text{RREF}(M)$. Then there exists a row of zeros in RREF(M), so $R^n(\lambda)\,\text{RREF}(M) = \text{RREF}(M)$ *for any* λ. Then:

$$\begin{aligned}
\det(MN) &= \det(E_1 E_2 \cdots E_k \text{ RREF}(M) N) \\
&= \det(E_1) \cdots \det(E_k) \det(\text{RREF}(M) N) \\
&= \det(E_1) \cdots \det(E_k) \det(R^n(\lambda) \text{ RREF}(M) N) \\
&= \det(E_1) \cdots \det(E_k) \lambda \det(\text{RREF}(M) N) \\
&= \lambda \det(MN)
\end{aligned}$$



Figure 8.6: "The determinant of a product is the product of determinants."

Which implies that $\det(MN) = 0 = \det M \det N$.

Thus we have shown that for *any* matrices M and N,

$$\det(MN) = \det M \det N$$

This result is *extremely important*; do not forget it!

Alternative proof

Reading homework: problem 4

8.3 Review Problems

Webwork:	Reading Problems	1, 2, 3, 4
	2×2 Determinant	7
	Determinants and invertibility	8, 9, 10, 11

1. Let

$$M = \begin{pmatrix} m^1_1 & m^1_2 & m^1_3 \\ m^2_1 & m^2_2 & m^2_3 \\ m^3_1 & m^3_2 & m^3_3 \end{pmatrix} .$$

Use row operations to put M into *row echelon form*. For simplicity, assume that $m_1^1 \neq 0 \neq m_1^1 m_2^2 - m_1^2 m_2^1$.

Prove that M is non-singular if and only if:

$$m_1^1 m_2^2 m_3^3 - m_1^1 m_3^2 m_2^3 + m_2^1 m_3^2 m_1^3 - m_2^1 m_1^2 m_3^3 + m_3^1 m_1^2 m_2^3 - m_3^1 m_2^2 m_1^3 \neq 0$$

2. (a) What does the matrix $E_2^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ do to $M = \begin{pmatrix} a & b \\ d & c \end{pmatrix}$ under left multiplication? What about right multiplication?
   (b) Find elementary matrices $R^1(\lambda)$ and $R^2(\lambda)$ that respectively multiply rows 1 and 2 of M by λ but otherwise leave M the same under left multiplication.
   (c) Find a matrix $S_2^1(\lambda)$ that adds a multiple λ of row 2 to row 1 under left multiplication.

3. Let $\hat{\sigma}$ denote the permutation obtained from σ by transposing the first two outputs, *i.e.* $\hat{\sigma}(1) = \sigma(2)$ and $\hat{\sigma}(2) = \sigma(1)$. Suppose the function $f : \{1, 2, 3, 4\} \to \mathbb{R}$. Write out explicitly the following two sums:
$$\sum_{\sigma} f\big(\sigma(s)\big) \text{ and } \sum_{\sigma} f\big(\hat{\sigma}(s)\big) .$$
What do you observe? Now write a brief explanation why the following equality holds
$$\sum_{\sigma} F(\sigma) = \sum_{\sigma} F(\hat{\sigma}) ,$$
where the domain of the function F is the set of all permutations of n objects and $\hat{\sigma}$ is related to σ by swapping a given pair of objects.

4. Let M be a matrix and $S_j^i M$ the same matrix with rows i and j switched. Explain every line of the series of equations proving that $\det M = -\det(S_j^i M)$.

5. Let M' be the matrix obtained from M by swapping two columns i and j. Show that $\det M' = -\det M$.

6. The scalar triple product of three vectors u, v, w from $\mathbb{R}^3$ is $u \cdot (v \times w)$. Show that this product is the same as the determinant of the matrix whose columns are u, v, w (in that order). What happens to the scalar triple product when the factors are permuted?

7. Show that if M is a 3×3 matrix whose third row is a sum of multiples of the other rows ($R_3 = aR_2 + bR_1$) then $\det M = 0$. Show that the same is true if one of the columns is a sum of multiples of the others.

8. Calculate the determinant below by factoring the matrix into elementary matrices times simpler matrices and using the trick

$$\det(M) = \det(E^{-1}EM) = \det(E^{-1})\det(EM).$$

Explicitly show each ERO matrix.

$$\det \begin{pmatrix} 2 & 1 & 0 \\ 4 & 3 & 1 \\ 2 & 2 & 2 \end{pmatrix}$$

9. Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $N = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$. Compute the following:

   (a) $\det M$.

   (b) $\det N$.

   (c) $\det(MN)$.

   (d) $\det M \det N$.

   (e) $\det(M^{-1})$ assuming $ad - bc \neq 0$.

   (f) $\det(M^T)$

   (g) $\det(M + N) - (\det M + \det N)$. Is the determinant a linear transformation from square matrices to real numbers? Explain.

10. Suppose $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible. Write M as a product of elementary row matrices times RREF(M).

11. Find the inverses of each of the elementary matrices, $E^i_j, R^i(\lambda), S^i_j(\lambda)$. Make sure to show that the elementary matrix times its inverse is actually the identity.

12. Let e^i_j denote the matrix with a 1 in the i-th row and j-th column and 0's everywhere else, and let A be an arbitrary 2×2 matrix. Compute $\det(A + tI_2)$. What is the first order term (the t^1 term)? Can you

express your results in terms of $\operatorname{tr}(A)$? What about the first order term in $\det(A + tI_n)$ for any arbitrary $n \times n$ matrix A in terms of $\operatorname{tr}(A)$?

Note that the result of $\det(A + tI_2)$ is a polynomial in the variable t known as the *characteristic polynomial*.

13. (Directional) Derivative of the determinant:
Notice that $\det \colon \mathbb{M}_n^n \to \mathbb{R}$ (where $\mathbb{M}_n^n$ is the vector space of all $n \times n$ matrices) det is a function of n^2 variables so we can take directional derivatives of det.

    Let A be an arbitrary $n \times n$ matrix, and for all i and j compute the following:

    (a)
    $$\lim_{t \to 0} \frac{\det(I_2 + te^i_j) - \det(I_2)}{t}$$

    (b)
    $$\lim_{t \to 0} \frac{\det(I_3 + te^i_j) - \det(I_3)}{t}$$

    (c)
    $$\lim_{t \to 0} \frac{\det(I_n + te^i_j) - \det(I_n)}{t}$$

    (d)
    $$\lim_{t \to 0} \frac{\det(I_n + At) - \det(I_n)}{t}$$

    Note, these are the directional derivative in the e^i_j and A directions.

14. How many functions are in the set
$$\{f : \{1, \ldots, n\} \to \{1, \ldots, n\} | f^{-1} \text{ exists}\}\,?$$

    What about the set
    $$\{1, \ldots, n\}^{\{1,\ldots,n\}}\,?$$

    Which of these two sets correspond to the set of all permutations of n objects?

8.4 Properties of the Determinant

We now know that the determinant of a matrix is non-zero if and only if that matrix is invertible. We also know that the determinant is a *multiplicative* function, in the sense that $\det(MN) = \det M \det N$. Now we will devise some methods for calculating the determinant.

Recall that:

$$\det M = \sum_{\sigma} \operatorname{sgn}(\sigma) m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)}.$$

A *minor* of an $n \times n$ matrix M is the determinant of any square matrix obtained from M by deleting one row and one column. In particular, any entry m^i_j of a square matrix M is associated to a minor obtained by deleting the ith row and jth column of M.

It is possible to write the determinant of a matrix in terms of its minors as follows:

$$\begin{aligned}
\det M &= \sum_{\sigma} \operatorname{sgn}(\sigma)\, m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)} \\
&= m^1_1 \sum_{\not\sigma^1} \operatorname{sgn}(\not\sigma^1)\, m^2_{\not\sigma^1(2)} \cdots m^n_{\not\sigma^1(n)} \\
&+ m^1_2 \sum_{\not\sigma^2} \operatorname{sgn}(\not\sigma^2)\, m^2_{\not\sigma^2(1)} m^3_{\not\sigma^2(3)} \cdots m^n_{\not\sigma^2(n)} \\
&+ m^1_3 \sum_{\not\sigma^3} \operatorname{sgn}(\not\sigma^3)\, m^2_{\not\sigma^3(1)} m^3_{\not\sigma^3(2)} m^4_{\not\sigma^3(4)} \cdots m^n_{\not\sigma^3(n)} \\
&+ \cdots
\end{aligned}$$

Here the symbols $\not\sigma^k$ refers to the permutation σ with the input k removed. The summand on the j'th line of the above formula looks like the determinant of the minor obtained by removing the first and j'th column of M. However we still need to replace sum of $\not\sigma^j$ by a sum over permutations of column numbers of the matrix entries of this minor. This costs a minus sign whenever $j-1$ is odd. In other words, to expand by minors we pick an entry m^1_j of the first row, then add $(-1)^{j-1}$ times the determinant of the matrix with row i and column j deleted. An example will probably help:

Example 103 Let's compute the determinant of

$$M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

using expansion by minors:

$$\begin{aligned} \det M &= 1 \det \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} - 2 \det \begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} + 3 \det \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} \\ &= 1(5 \cdot 9 - 8 \cdot 6) - 2(4 \cdot 9 - 7 \cdot 6) + 3(4 \cdot 8 - 7 \cdot 5) \\ &= 0 \end{aligned}$$

Here, M^{-1} does not exist because[1] $\det M = 0$.

Example 104 Sometimes the entries of a matrix allow us to simplify the calculation of the determinant. Take $N = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 0 & 0 \\ 7 & 8 & 9 \end{pmatrix}$. Notice that the second row has many zeros; then we can switch the first and second rows of N before expanding in minors to get:

$$\begin{aligned} \det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 0 & 0 \\ 7 & 8 & 9 \end{pmatrix} &= -\det \begin{pmatrix} 4 & 0 & 0 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{pmatrix} \\ &= -4 \det \begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix} \\ &= 24 \end{aligned}$$



`Example`



Since we know how the determinant of a matrix changes when you perform row operations, it is often very beneficial to perform row operations before computing the determinant by brute force.

[1]A fun exercise is to compute the determinant of a 4×4 matrix filled in order, from left to right, with the numbers $1, 2, 3, \ldots, 16$. What do you observe? Try the same for a 5×5 matrix with $1, 2, 3, \ldots, 25$. Is there a pattern? Can you explain it?

Example 105

$$\det\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}=\det\begin{pmatrix}1&2&3\\3&3&3\\6&6&6\end{pmatrix}=\det\begin{pmatrix}1&2&3\\3&3&3\\0&0&0\end{pmatrix}=0\,.$$

Try to determine which row operations we made at each step of this computation.

You might suspect that determinants have similar properties with respect to columns as what applies to rows:

> If M is a square matrix then $\det M^T = \det M$.

Proof. By definition,

$$\det M = \sum_{\sigma} \operatorname{sgn}(\sigma) m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)}.$$

For any permutation σ, there is a unique inverse permutation σ^{-1} that undoes σ. If σ sends $i \to j$, then σ^{-1} sends $j \to i$. In the two-line notation for a permutation, this corresponds to just flipping the permutation over. For example, if $\sigma = \begin{bmatrix}1&2&3\\2&3&1\end{bmatrix}$, then we can find σ^{-1} by flipping the permutation and then putting the columns in order:

$$\sigma^{-1} = \begin{bmatrix}2&3&1\\1&2&3\end{bmatrix} = \begin{bmatrix}1&2&3\\3&1&2\end{bmatrix}\,.$$

Since any permutation can be built up by transpositions, one can also find the inverse of a permutation σ by undoing each of the transpositions used to build up σ; this shows that one can use the same number of transpositions to build σ and σ^{-1}. In particular, $\operatorname{sgn}\sigma = \operatorname{sgn}\sigma^{-1}$.

Reading homework: problem 5

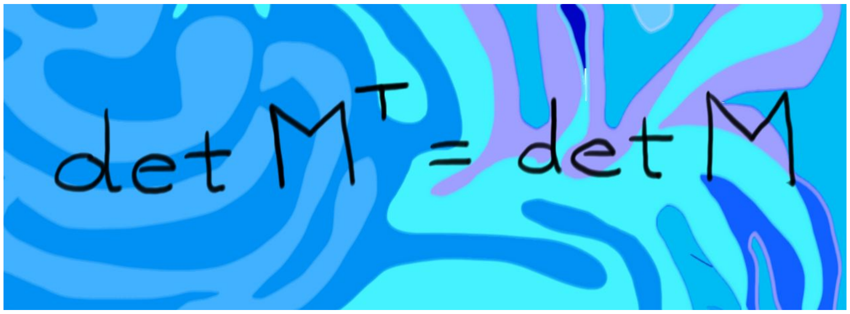


Figure 8.7: Transposes leave the determinant unchanged.

Then we can write out the above in formulas as follows:

$$\begin{aligned}
\det M &= \sum_\sigma \operatorname{sgn}(\sigma) m^1_{\sigma(1)} m^2_{\sigma(2)} \cdots m^n_{\sigma(n)} \\
&= \sum_\sigma \operatorname{sgn}(\sigma) m_1^{\sigma^{-1}(1)} m_2^{\sigma^{-1}(2)} \cdots m_n^{\sigma^{-1}(n)} \\
&= \sum_\sigma \operatorname{sgn}(\sigma^{-1}) m_1^{\sigma^{-1}(1)} m_2^{\sigma^{-1}(2)} \cdots m_n^{\sigma^{-1}(n)} \\
&= \sum_\sigma \operatorname{sgn}(\sigma) m_1^{\sigma(1)} m_2^{\sigma(2)} \cdots m_n^{\sigma(n)} \\
&= \det M^T .
\end{aligned}$$

The second-to-last equality is due to the existence of a unique inverse permutation: summing over permutations is the same as summing over all inverses of permutations (see review problem 3). The final equality is by the definition of the transpose. □

Example 106 Because of this, we see that expansion by minors also works over columns. Let

$$M = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 8 & 9 \end{pmatrix} .$$

Then

$$\det M = \det M^T = 1 \det \begin{pmatrix} 5 & 8 \\ 6 & 9 \end{pmatrix} = -3 .$$

8.4.1 Determinant of the Inverse

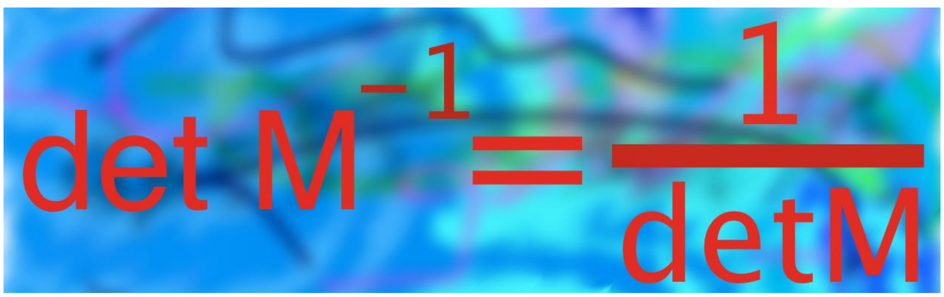
Let M and N be $n \times n$ matrices. We previously showed that

$$\det(MN) = \det M \det N, \text{ and } \det I = 1.$$

Then $1 = \det I = \det(MM^{-1}) = \det M \det M^{-1}$. As such we have:

Theorem 8.4.1.

$$\det M^{-1} = \frac{1}{\det M}$$



8.4.2 Adjoint of a Matrix

Recall that for a 2×2 matrix

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} I\,.$$

Or in a more careful notation: if

$$M = \begin{pmatrix} m_1^1 & m_2^1 \\ m_1^2 & m_2^2 \end{pmatrix},$$

then

$$M^{-1} = \frac{1}{m_1^1 m_2^2 - m_2^1 m_1^2} \begin{pmatrix} m_2^2 & -m_2^1 \\ -m_1^2 & m_1^1 \end{pmatrix},$$

so long as $\det M = m_1^1 m_2^2 - m_2^1 m_1^2 \neq 0$. The matrix $\begin{pmatrix} m_2^2 & -m_2^1 \\ -m_1^2 & m_1^1 \end{pmatrix}$ that appears above is a special matrix, called the *adjoint* of M. Let's define the adjoint for an $n \times n$ matrix.

The **cofactor** of M corresponding to the entry m_j^i of M is the product of the minor associated to m_j^i and $(-1)^{i+j}$. This is written cofactor(m_j^i).

Definition For $M = (m^i_j)$ a square matrix, the **adjoint matrix** $\operatorname{adj} M$ is given by

$$\operatorname{adj} M = (\operatorname{cofactor}(m^i_j))^T.$$

Example 107

$$\operatorname{adj}\begin{pmatrix}3 & -1 & -1\\ 1 & 2 & 0\\ 0 & 1 & 1\end{pmatrix} = \begin{pmatrix} \det\begin{pmatrix}2 & 0\\ 1 & 1\end{pmatrix} & -\det\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} & \det\begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix} \\ -\det\begin{pmatrix}-1 & -1\\ 1 & 1\end{pmatrix} & \det\begin{pmatrix}3 & -1\\ 0 & 1\end{pmatrix} & -\det\begin{pmatrix}3 & -1\\ 0 & 1\end{pmatrix} \\ \det\begin{pmatrix}-1 & -1\\ 2 & 0\end{pmatrix} & -\det\begin{pmatrix}3 & -1\\ 1 & 0\end{pmatrix} & \det\begin{pmatrix}3 & -1\\ 1 & 2\end{pmatrix}\end{pmatrix}^T$$

Reading homework: problem 6

Let's compute the product $M \operatorname{adj} M$. For any matrix N, the i, j entry of MN is given by taking the dot product of the ith row of M and the jth column of N. Notice that the dot product of the ith row of M and the ith column of $\operatorname{adj} M$ is just the expansion by minors of $\det M$ in the ith row. Further, notice that the dot product of the ith row of M and the jth column of $\operatorname{adj} M$ with $j \neq i$ is the same as expanding M by minors, but with the jth row replaced by the ith row. Since the determinant of any matrix with a row repeated is zero, then these dot products are zero as well.

We know that the i, j entry of the product of two matrices is the dot product of the ith row of the first by the jth column of the second. Then:

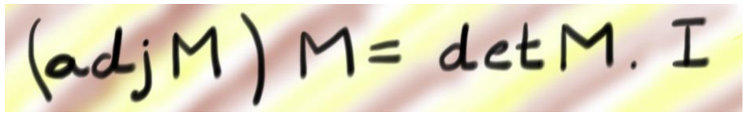
$$M \operatorname{adj} M = (\det M)I$$

Thus, when $\det M \neq 0$, the adjoint gives an explicit formula for M^{-1}.

Theorem 8.4.2. *For M a square matrix with $\det M \neq 0$ (equivalently, if M is invertible), then*

$$M^{-1} = \frac{1}{\det M} \operatorname{adj} M$$

`The Adjoint Matrix`



Example 108 Continuing with the previous example,

$$\operatorname{adj}\begin{pmatrix}3 & -1 & -1\\ 1 & 2 & 0\\ 0 & 1 & 1\end{pmatrix} = \begin{pmatrix}2 & 0 & 2\\ -1 & 3 & -1\\ 1 & -3 & 7\end{pmatrix}.$$

Now, multiply:

$$\begin{pmatrix}3 & -1 & -1\\ 1 & 2 & 0\\ 0 & 1 & 1\end{pmatrix}\begin{pmatrix}2 & 0 & 2\\ -1 & 3 & -1\\ 1 & -3 & 7\end{pmatrix} = \begin{pmatrix}6 & 0 & 0\\ 0 & 6 & 0\\ 0 & 0 & 6\end{pmatrix}$$

$$\Rightarrow \begin{pmatrix}3 & -1 & -1\\ 1 & 2 & 0\\ 0 & 1 & 1\end{pmatrix}^{-1} = \frac{1}{6}\begin{pmatrix}2 & 0 & 2\\ -1 & 3 & -1\\ 1 & -3 & 7\end{pmatrix}$$

This process for finding the inverse matrix is sometimes called *Cramer's Rule* .

8.4.3 Application: Volume of a Parallelepiped

Given three vectors u, v, w in $\mathbb{R}^3$, the parallelepiped determined by the three vectors is the "squished" box whose edges are parallel to u, v, and w as depicted in Figure 8.8.

You probably learnt in a calculus course that the volume of this object is $|u \cdot (v \times w)|$. This is the same as expansion by minors of the matrix whose columns are u, v, w. Then:

$$\text{Volume} = \big|\det\begin{pmatrix}u & v & w\end{pmatrix}\big|$$

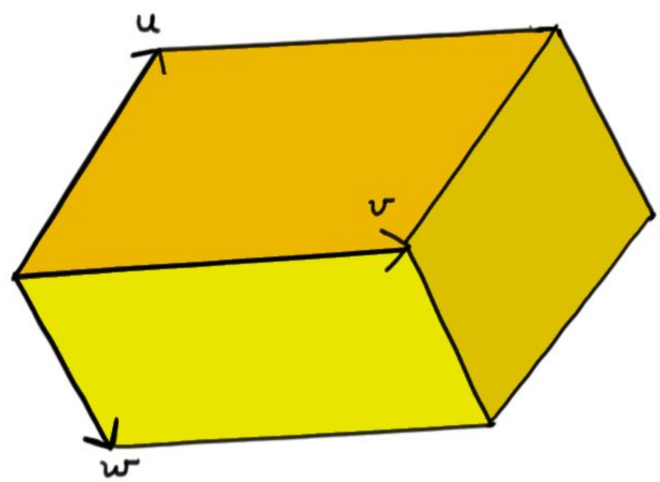


Figure 8.8: A parallelepiped.

8.5 Review Problems

Webwork:

Reading Problems	5,6
Row of zeros	12
3×3 determinant	13
Triangular determinants	14,15,16,17
Expanding in a column	18
Minors and cofactors	19

1. Find the determinant via expanding by minors.

$$\begin{pmatrix} 2 & 1 & 3 & 7 \\ 6 & 1 & 4 & 4 \\ 2 & 1 & 8 & 0 \\ 1 & 0 & 2 & 0 \end{pmatrix}$$

2. Even if M is not a square matrix, both MM^T and M^TM are square. Is it true that $\det(MM^T) = \det(M^TM)$ for all matrices M? How about $\operatorname{tr}(MM^T) = \operatorname{tr}(M^TM)$?

3. Let σ^{-1} denote the inverse permutation of σ. Suppose the function $f : \{1,2,3,4\} \to \mathbb{R}$. Write out explicitly the following two sums:

$$\sum_{\sigma} f\big(\sigma(s)\big) \text{ and } \sum_{\sigma} f\big(\sigma^{-1}(s)\big)\,.$$

What do you observe? Now write a brief explanation why the following equality holds

$$\sum_{\sigma} F(\sigma) = \sum_{\sigma} F(\sigma^{-1})\,,$$

where the domain of the function F is the set of all permutations of n objects.

4. Suppose $M = LU$ is an LU decomposition. Explain how you would efficiently compute $\det M$ in this case. How does this decomposition allow you to easily see if M is invertible?

5. In computer science, the *complexity* of an algorithm is (roughly) computed by counting the number of times a given operation is performed. Suppose adding or subtracting any two numbers takes a seconds, and multiplying two numbers takes m seconds. Then, for example, computing $2 \cdot 6 - 5$ would take $a + m$ seconds.

   (a) How many additions and multiplications does it take to compute the determinant of a general 2×2 matrix?

   (b) Write a formula for the number of additions and multiplications it takes to compute the determinant of a general $n \times n$ matrix using the definition of the determinant as a sum over permutations. Assume that finding and multiplying by the sign of a permutation is free.

   (c) How many additions and multiplications does it take to compute the determinant of a general 3×3 matrix using expansion by minors? Assuming $m = 2a$, is this faster than computing the determinant from the definition?



Hint



9

Subspaces and Spanning Sets

It is time to study vector spaces more carefully and return to some fundamental questions:

1. *Subspaces*: When is a subset of a vector space itself a vector space? (This is the notion of a *subspace*.)

2. *Linear Independence*: Given a collection of vectors, is there a way to tell whether they are independent, or if one is a "linear combination" of the others?

3. *Dimension*: Is there a consistent definition of how "big" a vector space is?

4. *Basis*: How do we label vectors? Can we write any vector as a sum of some basic set of vectors? How do we change our point of view from vectors labeled one way to vectors labeled in another way?

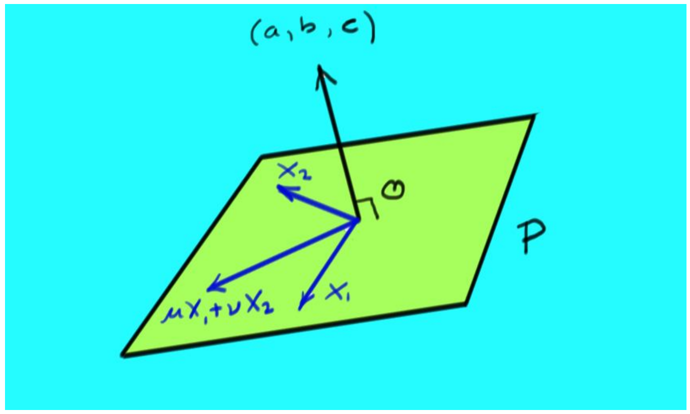
Let's start at the top!

9.1 Subspaces

Definition We say that a subset U of a vector space V is a **subspace** of V if U is a vector space under the inherited addition and scalar multiplication operations of V.

Example 109 Consider a plane P in $\mathbb{R}^3$ through the origin:

$$ax + by + cz = 0.$$



This equation can be expressed as the homogeneous system $\begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$, or $MX = 0$ with M the matrix $\begin{pmatrix} a & b & c \end{pmatrix}$. If X_1 and X_2 are both solutions to $MX = 0$, then, by linearity of matrix multiplication, so is $\mu X_1 + \nu X_2$:

$$M(\mu X_1 + \nu X_2) = \mu M X_1 + \nu M X_2 = 0.$$

So P is closed under addition and scalar multiplication. Additionally, P contains the origin (which can be derived from the above by setting $\mu = \nu = 0$). All other vector space requirements hold for P because they hold for all vectors in $\mathbb{R}^3$.

Theorem 9.1.1 (Subspace Theorem)**.** *Let U be a non-empty subset of a vector space V. Then U is a subspace if and only if $\mu u_1 + \nu u_2 \in U$ for arbitrary u_1, u_2 in U, and arbitrary constants μ, ν.*

Proof. One direction of this proof is easy: if U is a subspace, then it is a vector space, and so by the additive closure and multiplicative closure properties of vector spaces, it has to be true that $\mu u_1 + \nu u_2 \in U$ for all u_1, u_2 in U and all constants constants μ, ν.

The other direction is almost as easy: we need to show that if $\mu u_1 + \nu u_2 \in U$ for all u_1, u_2 in U and all constants μ, ν, then U is a vector space. That is, we need to show that the ten properties of vector spaces are satisfied. We already know that the additive closure and multiplicative closure properties are satisfied. Further, U has all of the other eight properties because V has them. □

Note that the requirements of the subspace theorem are often referred to as "closure".

We can use this theorem to check if a set is a vector space. That is, if we have some set U of vectors that come from some bigger vector space V, to check if U itself forms a smaller vector space we need check only two things:

1. If we add any two vectors in U, do we end up with a vector in U?
2. If we multiply any vector in U by any constant, do we end up with a vector in U?

If the answer to both of these questions is yes, then U is a vector space. If not, U is not a vector space.


Reading homework: problem 1

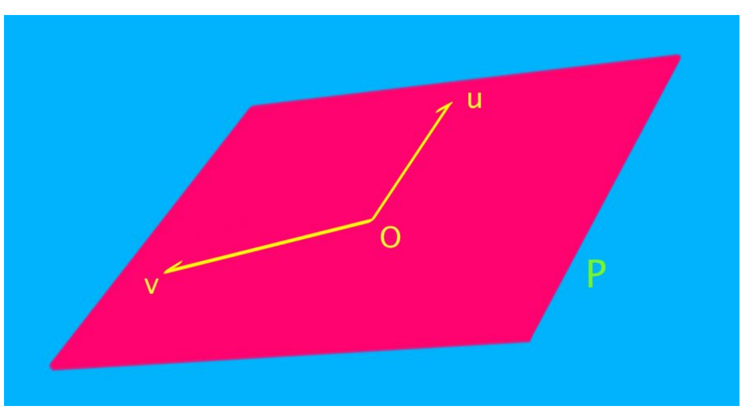
9.2 Building Subspaces

Consider the set

$$U = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \subset \mathbb{R}^3 .$$

Because U consists of only two vectors, it clear that U is *not* a vector space, since any constant multiple of these vectors should also be in U. For example, the 0-vector is not in U, nor is U closed under vector addition.

But we know that any two vectors define a plane:



In this case, the vectors in U define the xy-plane in $\mathbb{R}^3$. We can view the xy-plane as the set of all vectors that arise as a linear combination of the two vectors in U. We call this set of all linear combinations the *span* of U:

$$\operatorname{span}(U) = \left\{ x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \middle| x, y \in \mathbb{R} \right\}.$$

Notice that any vector in the xy-plane is of the form

$$\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \in \operatorname{span}(U).$$

Definition Let V be a vector space and $S = \{s_1, s_2, \ldots\} \subset V$ a subset of V. Then the **span of** S, denoted $\operatorname{span}(S)$, is the set

$$\operatorname{span}(S) := \{r^1 s_1 + r^2 s_2 + \cdots + r^N s_N \mid r^i \in \mathbb{R}, N \in \mathbb{N}\}.$$

That is, the span of S is the set of all finite linear combinations[1] of elements of S. Any *finite* sum of the form "a constant times s_1 plus a constant times s_2 plus a constant times s_3 and so on" is in the span of S.[2].

Example 110 Let $V = \mathbb{R}^3$ and $X \subset V$ be the x-axis. Let $P = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and set

$$S = X \cup \{P\}.$$

The vector $\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$ is in $\operatorname{span}(S)$, because $\begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Similarly, the vector $\begin{pmatrix} -12 \\ 17.5 \\ 0 \end{pmatrix}$ is in $\operatorname{span}(S)$, because $\begin{pmatrix} -12 \\ 17.5 \\ 0 \end{pmatrix} = \begin{pmatrix} -12 \\ 0 \\ 0 \end{pmatrix} + 17.5 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Similarly, any vector

[1] Usually our vector spaces are defined over $\mathbb{R}$, but in general we can have vector spaces defined over different base fields such as $\mathbb{C}$ or $\mathbb{Z}_2$. The coefficients r^i should come from whatever our base field is (usually $\mathbb{R}$).

[2] It is important that we only allow finitely many terms in our linear combinations; in the definition above, N must be a finite number. It can be any finite number, but it must be finite. We can relax the requirement that $S = \{s_1, s_2, \ldots\}$ and just let S be any set of vectors. Then we shall write $\operatorname{span}(S) := \{r^1 s_1 + r^2 s_2 + \cdots + r^N s_N \mid r^i \in \mathbb{R}, s_i \in S, N \in \mathbb{N}, \}$

of the form

$$\begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

is in $\operatorname{span}(S)$. On the other hand, any vector in $\operatorname{span}(S)$ must have a zero in the z-coordinate. (Why?) So $\operatorname{span}(S)$ is the xy-plane, which is a vector space. (Try drawing a picture to verify this!)

Reading homework: problem 2

Lemma 9.2.1. *For any subset $S \subset V$, $\operatorname{span}(S)$ is a subspace of V.*

Proof. We need to show that $\operatorname{span}(S)$ is a vector space.

It suffices to show that $\operatorname{span}(S)$ is closed under linear combinations. Let $u, v \in \operatorname{span}(S)$ and λ, μ be constants. By the definition of $\operatorname{span}(S)$, there are constants c^i and d^i (some of which could be zero) such that:

$$\begin{aligned} u &= c^1 s_1 + c^2 s_2 + \cdots \\ v &= d^1 s_1 + d^2 s_2 + \cdots \\ \Rightarrow \lambda u + \mu v &= \lambda(c^1 s_1 + c^2 s_2 + \cdots) + \mu(d^1 s_1 + d^2 s_2 + \cdots) \\ &= (\lambda c^1 + \mu d^1) s_1 + (\lambda c^2 + \mu d^2) s_2 + \cdots \end{aligned}$$

This last sum is a linear combination of elements of S, and is thus in $\operatorname{span}(S)$. Then $\operatorname{span}(S)$ is closed under linear combinations, and is thus a subspace of V. □

Note that this proof, like many proofs, consisted of little more than just writing out the definitions.

Example 111 For which values of a does

$$\operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ a \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}, \begin{pmatrix} a \\ 1 \\ 0 \end{pmatrix} \right\} = \mathbb{R}^3 ?$$

Given an arbitrary vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ in $\mathbb{R}^3$, we need to find constants r^1, r^2, r^3 such that

$$r^1 \begin{pmatrix} 1 \\ 0 \\ a \end{pmatrix} + r^2 \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} + r^3 \begin{pmatrix} a \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

We can write this as a linear system in the unknowns r^1, r^2, r^3 as follows:

$$\begin{pmatrix} 1 & 1 & a \\ 0 & 2 & 1 \\ a & -3 & 0 \end{pmatrix} \begin{pmatrix} r^1 \\ r^2 \\ r^3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

If the matrix $M = \begin{pmatrix} 1 & 1 & a \\ 0 & 2 & 1 \\ a & -3 & 0 \end{pmatrix}$ is invertible, then we can find a solution

$$M^{-1} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r^1 \\ r^2 \\ r^3 \end{pmatrix}$$

for *any* vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$.

Therefore we should choose a so that M is invertible:

$$i.e., \; 0 \neq \det M = -2a^2 + 3 + a = -(2a-3)(a+1).$$

Then the span is $\mathbb{R}^3$ if and only if $a \neq -1, \frac{3}{2}$.

Linear systems as spanning sets

Some other very important ways of building subspaces are given in the following examples.

Example 112 (The kernel of a linear map).

Suppose $L : U \to V$ is a linear map between vector spaces. Then if

$$L(u) = 0 = L(u') ,$$

linearity tells us that

$$L(\alpha u + \beta u') = \alpha L(u) + \beta L(u') = \alpha 0 + \beta 0 = 0 .$$

Hence, thanks to the subspace theorem, the set of all vectors in U that are mapped to the zero vector is a subspace of V. It is called the kernel of L:

$$\ker L := \{u \in U | L(u) = 0\} \subset U.$$

Note that finding a kernel means finding a solution to a homogeneous linear equation.

Example 113 (The image of a linear map).

Suppose $L : U \to V$ is a linear map between vector spaces. Then if

$$v = L(u) \text{ and } v' = L(u'),$$

linearity tells us that

$$\alpha v + \beta v' = \alpha L(u) + \beta L(u') = L(\alpha u + \beta u').$$

Hence, calling once again on the subspace theorem, the set of all vectors in V that are obtained as outputs of the map L is a subspace. It is called the image of L:

$$\text{im} L := \{L(u) \mid u \in U\} \subset V.$$

Example 114 (An eigenspace of a linear map).

Suppose $L : V \to V$ is a linear map and V is a vector space. Then if

$$L(u) = \lambda u \text{ and } L(v) = \lambda v,$$

linearity tells us that

$$L(\alpha u + \beta v) = \alpha L(u) + \beta L(v) = \alpha L(u) + \beta L(v) = \alpha \lambda u + \beta \lambda v = \lambda(\alpha u + \beta v).$$

Hence, again by subspace theorem, the set of all vectors in V that obey the *eigenvector equation* $L(v) = \lambda v$ is a subspace of V. It is called an eigenspace

$$V_\lambda := \{v \in V | L(v) = \lambda v\}.$$

For most scalars λ, the only solution to $L(v) = \lambda v$ will be $v = 0$, which yields the trivial subspace $\{0\}$. When there are nontrivial solutions to $L(v) = \lambda v$, the number λ is called an eigenvalue, and carries essential information about the map L.

Kernels, images and eigenspaces are discussed in great depth in chapters 16 and 12.

9.3 Review Problems

Webwork:

Reading Problems	1, 2
Subspaces	3, 4, 5, 6
Spans	7, 8

1. Determine if $x - x^3 \in \text{span}\{x^2, 2x + x^2, x + x^3\}$.

2. Let U and W be subspaces of V. Are:

   (a) $U \cup W$

   (b) $U \cap W$

   also subspaces? Explain why or why not. Draw examples in $\mathbb{R}^3$.


Hint


3. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ where

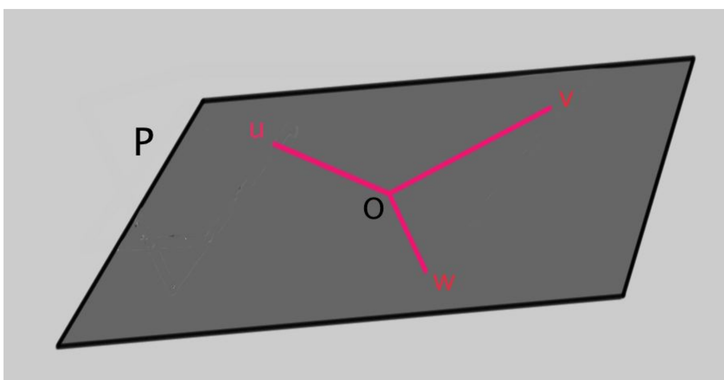
$$L(x, y, z) = (x + 2y + z, 2x + y + z, 0)\,.$$

   Find $\ker L$, $\text{im} L$ and the eigenspaces $\mathbb{R}^3_{-1}$, $\mathbb{R}^3_{3}$. Your answers should be subsets of $\mathbb{R}^3$. Express them using span notation.

10

Linear Independence

Consider a plane P that includes the origin in $\mathbb{R}^3$ and non-zero vectors $\{u, v, w\}$ in P.



If no two of u, v and w are parallel, then $P = \text{span}\{u, v, w\}$. But any two vectors determines a plane, so we should be able to span the plane using only two of the vectors u, v, w. Then we could choose two of the vectors in $\{u, v, w\}$ whose span is P, and express the other as a linear combination of those two. Suppose u and v span P. Then there exist constants d^1, d^2 (not both zero) such that $w = d^1 u + d^2 v$. Since w can be expressed in terms of u and v we say that it is not independent. More generally, the relationship

$$c^1 u + c^2 v + c^3 w = 0 \qquad c^i \in \mathbb{R}, \text{ some } c^i \neq 0$$

expresses the fact that u, v, w are not all independent.

Definition We say that the vectors $v_1, v_2, \dots, v_n$ are **linearly dependent** if there exist constants[1] $c^1, c^2, \dots, c^n$ not all zero such that

$$c^1 v_1 + c^2 v_2 + \cdots + c^n v_n = 0.$$

Otherwise, the vectors $v_1, v_2, \dots, v_n$ are **linearly independent.**

Remark The zero vector 0_V can *never* be on a list of independent vectors because $\alpha 0_V = 0_V$ for any scalar α.

Example 115 Consider the following vectors in $\mathbb{R}^3$:

$$v_1 = \begin{pmatrix} 4 \\ -1 \\ 3 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} -3 \\ 7 \\ 4 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 5 \\ 12 \\ 17 \end{pmatrix}, \qquad v_4 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Are these vectors linearly independent?

No, since $3v_1 + 2v_2 - v_3 + v_4 = 0$, the vectors are linearly *dependent*.



Worked Example



10.1 Showing Linear Dependence

In the above example we were given the linear combination $3v_1 + 2v_2 - v_3 + v_4$ seemingly by magic. The next example shows how to find such a linear combination, if it exists.

Example 116 Consider the following vectors in $\mathbb{R}^3$:

$$v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Are they linearly independent?

We need to see whether the system

$$c^1 v_1 + c^2 v_2 + c^3 v_3 = 0$$

[1] Usually our vector spaces are defined over $\mathbb{R}$, but in general we can have vector spaces defined over different base fields such as $\mathbb{C}$ or $\mathbb{Z}_2$. The coefficients c^i should come from whatever our base field is (usually $\mathbb{R}$).

has any solutions for c^1, c^2, c^3. We can rewrite this as a homogeneous system by building a matrix whose columns are the vectors v_1, v_2 and v_3:

$$\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \\ c^3 \end{pmatrix} = 0.$$

This system has solutions if and only if the matrix $M = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}$ is singular, so we should find the determinant of M:

$$\det M = \det \begin{pmatrix} 0 & 1 & 1 \\ 0 & 2 & 2 \\ 1 & 1 & 3 \end{pmatrix} = \det \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} = 0.$$

Therefore nontrivial solutions exist. At this point we know that the vectors are linearly dependent. If we need to, we can find coefficients that demonstrate linear dependence by solving

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 0 \\ 0 & 2 & 2 & 0 \\ 1 & 1 & 3 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 3 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The solution set $\{\mu(-2, -1, 1) \mid \mu \in \mathbb{R}\}$ encodes the linear combinations equal to zero; any choice of μ will produce coefficients c^1, c^2, c^3 that satisfy the linear homogeneous equation. In particular, $\mu = 1$ corresponds to the equation

$$c^1 v_1 + c^2 v_2 + c^3 v_3 = 0 \Rightarrow -2v_1 - v_2 + v_3 = 0.$$

Reading homework: problem 1

Definition Any sum of vectors $v_1, \ldots, v_k$ multiplied by scalars $c^1, \ldots, c^k$, namely

$$c^1 v_1 + \cdots + c^k v_k\,,$$

is called a *linear combination* of $v_1, \ldots, v_k$.

Theorem 10.1.1 (Linear Dependence)**.** *An ordered set of non-zero vectors $(v_1, \ldots, v_n)$ is linearly dependent if and only if one of the vectors v_k is expressible as a linear combination of the preceding vectors.*

Proof. The theorem is an if and only if statement, so there are two things to show.

i. First, we show that if $v_k = c^1 v_1 + \cdots c^{k-1} v_{k-1}$ then the set is linearly dependent.

This is easy. We just rewrite the assumption:

$$c^1 v_1 + \cdots + c^{k-1} v_{k-1} - v_k + 0 v_{k+1} + \cdots + 0 v_n = 0.$$

This is a vanishing linear combination of the vectors $\{v_1, \ldots, v_n\}$ with not all coefficients equal to zero, so $\{v_1, \ldots, v_n\}$ is a linearly dependent set.

ii. Now we show that linear dependence implies that there exists k for which v_k is a linear combination of the vectors $\{v_1, \ldots, v_{k-1}\}$.

The assumption says that

$$c^1 v_1 + c^2 v_2 + \cdots + c^n v_n = 0.$$

Take k to be the largest number for which c_k is not equal to zero. So:

$$c^1 v_1 + c^2 v_2 + \cdots + c^{k-1} v_{k-1} + c^k v_k = 0.$$

(Note that $k > 1$, since otherwise we would have $c^1 v_1 = 0 \Rightarrow v_1 = 0$, contradicting the assumption that none of the v_i are the zero vector.)

So we can rearrange the equation:

$$\begin{aligned} c^1 v_1 + c^2 v_2 + \cdots + c^{k-1} v_{k-1} &= -c^k v_k \\ \Rightarrow \ -\frac{c^1}{c^k} v_1 - \frac{c^2}{c^k} v_2 - \cdots - \frac{c^{k-1}}{c^k} v_{k-1} &= v_k. \end{aligned}$$

Therefore we have expressed v_k as a linear combination of the previous vectors, and we are done.

□



Worked proof



Example 117 Consider the vector space $P_2(t)$ of polynomials of degree less than or equal to 2. Set:

$$\begin{aligned} v_1 &= 1+t \\ v_2 &= 1+t^2 \\ v_3 &= t+t^2 \\ v_4 &= 2+t+t^2 \\ v_5 &= 1+t+t^2. \end{aligned}$$

The set $\{v_1, \ldots, v_5\}$ is linearly dependent, because $v_4 = v_1 + v_2$.

10.2 Showing Linear Independence

We have seen two different ways to show a set of vectors is linearly dependent: we can either find a linear combination of the vectors which is equal to zero, or we can express one of the vectors as a linear combination of the other vectors. On the other hand, to check that a set of vectors is linearly *independent*, we must check that every linear combination of our vectors with non-vanishing coefficients gives something other than the zero vector. Equivalently, to show that the set $v_1, v_2, \ldots, v_n$ is linearly independent, we must show that the equation $c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0$ has no solutions other than $c_1 = c_2 = \cdots = c_n = 0$.

Example 118 Consider the following vectors in $\mathbb{R}^3$:

$$v_1 = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix}.$$

Are they linearly independent?

We need to see whether the system

$$c^1v_1 + c^2v_2 + c^3v_3 = 0$$

has any solutions for c^1, c^2, c^3. We can rewrite this as a homogeneous system:

$$\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \\ c^3 \end{pmatrix} = 0.$$

This system has solutions if and only if the matrix $M = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}$ is singular, so we should find the determinant of M:

$$\det M = \det \begin{pmatrix} 0 & 2 & 1 \\ 0 & 2 & 4 \\ 2 & 1 & 3 \end{pmatrix} = 2 \det \begin{pmatrix} 2 & 1 \\ 2 & 4 \end{pmatrix} = 12.$$

Since the matrix M has non-zero determinant, the only solution to the system of equations

$$\begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \\ c^3 \end{pmatrix} = 0$$

is $c_1 = c_2 = c_3 = 0$. So the vectors v_1, v_2, v_3 are linearly independent.

Here is another example with bits:

Example 119 Let $\mathbb{Z}_2^3$ be the space of 3×1 bit-valued matrices (i.e., column vectors). Is the following subset linearly independent?

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$$

If the set is linearly dependent, then we can find non-zero solutions to the system:

$$c^1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c^2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + c^3 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 0,$$

which becomes the linear system

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \\ c^3 \end{pmatrix} = 0.$$

Solutions exist if and only if the determinant of the matrix is non-zero. But:

$$\det \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} = 1 \det \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} - 1 \det \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = -1 - 1 = 1 + 1 = 0$$

Therefore non-trivial solutions exist, and the set is not linearly independent.

Reading homework: problem 2

10.3 From Dependent Independent

Now suppose vectors $v_1, \ldots, v_n$ are linearly dependent,

$$c^1 v_1 + c^2 v_2 + \cdots + c^n v_n = 0$$

with $c^1 \neq 0$. Then:

$$\operatorname{span}\{v_1, \ldots, v_n\} = \operatorname{span}\{v_2, \ldots, v_n\}$$

because any $x \in \operatorname{span}\{v_1, \ldots, v_n\}$ is given by

$$\begin{aligned} x &= a^1 v_1 + \cdots + a^n v_n \\ &= a^1 \left(-\frac{c^2}{c_1} v_2 - \cdots - \frac{c^n}{c_1} v_n \right) + a^2 v_2 + \cdots + a^n v_n \\ &= \left(a^2 - a^1 \frac{c^2}{c_1} \right) v_2 + \cdots + \left(a^n - a^1 \frac{c^n}{c_1} \right) v_n . \end{aligned}$$

Then x is in $\operatorname{span}\{v_2, \ldots, v_n\}$.

When we write a vector space as the span of a list of vectors, we would like that list to be as short as possible (this idea is explored further in chapter 11). This can be achieved by iterating the above procedure.

Example 120 In the above example, we found that $v_4 = v_1 + v_2$. In this case, any expression for a vector as a linear combination involving v_4 can be turned into a combination without v_4 by making the substitution $v_4 = v_1 + v_2$.

Then:

$$\begin{aligned} S &= \operatorname{span}\{1+t, 1+t^2, t+t^2, 2+t+t^2, 1+t+t^2\} \\ &= \operatorname{span}\{1+t, 1+t^2, t+t^2, 1+t+t^2\}. \end{aligned}$$

Now we notice that $1+t+t^2 = \frac{1}{2}(1+t) + \frac{1}{2}(1+t^2) + \frac{1}{2}(t+t^2)$. So the vector $1+t+t^2 = v_5$ is also extraneous, since it can be expressed as a linear combination of the remaining three vectors, v_1, v_2, v_3. Therefore

$$S = \operatorname{span}\{1+t, 1+t^2, t+t^2\}.$$

In fact, you can check that there are no (non-zero) solutions to the linear system

$$c^1(1+t) + c^2(1+t^2) + c^3(t+t^2) = 0.$$

Therefore the remaining vectors $\{1+t, 1+t^2, t+t^2\}$ are linearly independent, and span the vector space S. Then these vectors are a minimal spanning set, in the sense that no more vectors can be removed since the vectors are linearly independent. Such a set is called a *basis* for S.

10.4 Review Problems

Webwork:

Reading Problems	1, 2
Testing for linear independence	3, 4
Gaussian elimination	5
Spanning and linear independence	6

1. Let B^n be the space of $n \times 1$ bit-valued matrices (*i.e.*, column vectors) over the field $\mathbb{Z}_2$. Remember that this means that the coefficients in any linear combination can be only 0 or 1, with rules for adding and multiplying coefficients given here.

   (a) How many different vectors are there in B^n?

   (b) Find a collection S of vectors that span B^3 and are linearly independent. In other words, find a basis of B^3.

   (c) Write each other vector in B^3 as a linear combination of the vectors in the set S that you chose.

   (d) Would it be possible to span B^3 with only two vectors?

   Hint

2. Let e_i be the vector in $\mathbb{R}^n$ with a 1 in the ith position and 0's in every other position. Let v be an arbitrary vector in $\mathbb{R}^n$.

   (a) Show that the collection $\{e_1, \ldots, e_n\}$ is linearly independent.

   (b) Demonstrate that $v = \sum_{i=1}^{n}(v \cdot e_i)e_i$.

   (c) The span$\{e_1, \ldots, e_n\}$ is the same as what vector space?

3. Consider the ordered set of vectors from $\mathbb{R}^3$

$$\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ 5 \end{pmatrix} \right)$$

   (a) Determine if the set is linearly independent by using the vectors as the columns of a matrix M and finding RREF(M).

(b) If possible, write each vector as a linear combination of the preceding ones.

(c) Remove the vectors which can be expressed as linear combinations of the preceding vectors to form a linearly independent ordered set. (Every vector in your set set should be from the given set.)

4. Gaussian elimination is a useful tool to figure out whether a set of vectors spans a vector space and if they are linearly independent. Consider a matrix M made from an ordered set of column vectors $(v_1, v_2, \ldots, v_m) \subset \mathbb{R}^n$ and the three cases listed below:

   (a) RREF(M) is the identity matrix.

   (b) RREF(M) has a row of zeros.

   (c) Neither case (a) or (b) apply.

   First give an explicit example for each case, state whether the column vectors you use are linearly independent or spanning in each case. Then, in general, determine whether $(v_1, v_2, \ldots, v_m)$ are linearly independent and/or spanning $\mathbb{R}^n$ in each of the three cases. If they are linearly dependent, does RREF(M) tell you which vectors could be removed to yield an independent set of vectors?

11 Basis and Dimension

In chapter 10, the notions of a linearly independent set of vectors in a vector space V, and of a set of vectors that span V were established; any set of vectors that span V can be reduced to some minimal collection of linearly independent vectors; such a minimal set is called a *basis* of the subspace V.

Definition Let V be a vector space. Then a set S is a **basis** for V if S is linearly independent and $V = \operatorname{span} S$.

If S is a basis of V and S has only finitely many elements, then we say that V is **finite-dimensional**. The number of vectors in S is the **dimension** of V.

Suppose V is a *finite-dimensional* vector space, and S and T are two different bases for V. One might worry that S and T have a different number of vectors; then we would have to talk about the dimension of V in terms of the basis S or in terms of the basis T. Luckily this isn't what happens. Later in this chapter, we will show that S and T must have the same number of vectors. This means that the dimension of a vector space is basis-independent. In fact, dimension is a very important characteristic of a vector space.

Example 121 $P_n(t)$ (polynomials in t of degree n or less) has a basis $\{1, t, \ldots, t^n\}$, since every vector in this space is a sum

$$a^0\,1 + a^1\,t + \cdots + a^n\,t^n, \qquad a^i \in \mathbb{R}\,,$$

so $P_n(t) = \operatorname{span}\{1, t, \ldots, t^n\}$. This set of vectors is linearly independent; If the polynomial $p(t) = c^0 1 + c^1 t + \cdots + c^n t^n = 0$, then $c^0 = c^1 = \cdots = c^n = 0$, so $p(t)$ is the zero polynomial. Thus $P_n(t)$ is finite dimensional, and $\dim P_n(t) = n + 1$.

Theorem 11.0.1. *Let $S = \{v_1, \ldots, v_n\}$ be a basis for a vector space V. Then every vector $w \in V$ can be written* uniquely *as a linear combination of vectors in the basis S:*

$$w = c^1 v_1 + \cdots + c^n v_n .$$

Proof. Since S is a basis for V, then $\operatorname{span} S = V$, and so there exist constants c^i such that $w = c^1 v_1 + \cdots + c^n v_n$.

Suppose there exists a second set of constants d^i such that

$$w = d^1 v_1 + \cdots + d^n v_n \, .$$

Then

$$\begin{aligned} 0_V &= w - w \\ &= c^1 v_1 + \cdots + c^n v_n - d^1 v_1 - \cdots - d^n v_n \\ &= (c^1 - d^1) v_1 + \cdots + (c^n - d^n) v_n . \end{aligned}$$

If it occurs exactly once that $c^i \neq d^i$, then the equation reduces to $0 = (c^i - d^i) v_i$, which is a contradiction since the vectors v_i are assumed to be non-zero.

If we have more than one i for which $c^i \neq d^i$, we can use this last equation to write one of the vectors in S as a linear combination of other vectors in S, which contradicts the assumption that S is linearly independent. Then for every i, $c^i = d^i$. □

`Proof Explanation`

Remark This theorem is the one that makes bases so useful–they allow us to convert abstract vectors into column vectors. By ordering the set S we obtain $B = (v_1, \ldots, v_n)$ and can write

$$w = (v_1, \ldots, v_n) \begin{pmatrix} c^1 \\ \vdots \\ c^n \end{pmatrix} = \begin{pmatrix} c^1 \\ \vdots \\ c^n \end{pmatrix}_B .$$

Remember that in general it makes no sense to drop the subscript B on the column vector on the right–most vector spaces are not made from columns of numbers!



Worked Example



Next, we would like to establish a method for determining whether a collection of vectors forms a basis for $\mathbb{R}^n$. But first, we need to show that any two bases for a finite-dimensional vector space has the same number of vectors.

Lemma 11.0.2. *If $S = \{v_1, \ldots, v_n\}$ is a basis for a vector space V and $T = \{w_1, \ldots, w_m\}$ is a linearly independent set of vectors in V, then $m \leq n$.*

The idea of the proof is to start with the set S and replace vectors in S one at a time with vectors from T, such that after each replacement we still have a basis for V.



Reading homework: problem 1

Proof. Since S spans V, then the set $\{w_1, v_1, \ldots, v_n\}$ is linearly dependent. Then we can write w_1 as a linear combination of the v_i; using that equation, we can express one of the v_i in terms of w_1 and the remaining v_j with $j \neq i$. Then we can discard one of the v_i from this set to obtain a linearly independent set that still spans V. Now we need to prove that S_1 is a basis; we must show that S_1 is linearly independent and that S_1 spans V.

The set $S_1 = \{w_1, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n\}$ is linearly independent: By the previous theorem, there was a unique way to express w_1 in terms of the set S. Now, to obtain a contradiction, suppose there is some k and constants c^i such that

$$v_k = c^0 w_1 + c^1 v_1 + \cdots + c^{i-1} v_{i-1} + c^{i+1} v_{i+1} + \cdots + c^n v_n.$$

Then replacing w_1 with its expression in terms of the collection S gives a way to express the vector v_k as a linear combination of the vectors in S, which contradicts the linear independence of S. On the other hand, we cannot express w_1 as a linear combination of the vectors in $\{v_j | j \neq i\}$, since the expression of w_1 in terms of S was unique, and had a non-zero coefficient for the vector v_i. Then no vector in S_1 can be expressed as a combination of other vectors in S_1, which demonstrates that S_1 is linearly independent.

The set S_1 spans V: For any $u \in V$, we can express u as a linear combination of vectors in S. But we can express v_i as a linear combination of

vectors in the collection S_1; rewriting v_i as such allows us to express u as a linear combination of the vectors in S_1. Thus S_1 is a basis of V with n vectors.

We can now iterate this process, replacing one of the v_i in S_1 with w_2, and so on. If $m \leq n$, this process ends with the set $S_m = \{w_1, \ldots, w_m, v_{i_1}, \ldots, v_{i_{n-m}}\}$, which is fine.

Otherwise, we have $m > n$, and the set $S_n = \{w_1, \ldots, w_n\}$ is a basis for V. But we still have some vector w_{n+1} in T that is not in S_n. Since S_n is a basis, we can write w_{n+1} as a combination of the vectors in S_n, which contradicts the linear independence of the set T. Then it must be the case that $m \leq n$, as desired. □



Worked Example



Corollary 11.0.3. *For a finite-dimensional vector space V, any two bases for V have the same number of vectors.*

Proof. Let S and T be two bases for V. Then both are linearly independent sets that span V. Suppose S has n vectors and T has m vectors. Then by the previous lemma, we have that $m \leq n$. But (exchanging the roles of S and T in application of the lemma) we also see that $n \leq m$. Then $m = n$, as desired. □



Reading homework: problem 2

11.1 Bases in $\mathbb{R}^n$.

In review question 2, chapter 10 you checked that

$$\mathbb{R}^n = \operatorname{span}\left\{\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \ldots, \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}\right\},$$

and that this set of vectors is linearly independent. (If you didn't do that problem, check this before reading any further!) So this set of vectors is

a basis for $\mathbb{R}^n$, and $\dim \mathbb{R}^n = n$. This basis is often called the *standard* or *canonical basis* for $\mathbb{R}^n$. The vector with a one in the ith position and zeros everywhere else is written e_i. (You could also view it as the function $\{1, 2, \ldots, n\} \to \mathbb{R}$ where $e_i(j) = 1$ if $i = j$ and 0 if $i \neq j$.) It points in the direction of the ith coordinate axis, and has unit length. In multivariable calculus classes, this basis is often written $\{\hat{\imath}, \hat{\jmath}, \hat{k}\}$ for $\mathbb{R}^3$.

Note that it is often convenient to order basis elements, so rather than writing a set of vectors, we would write a list. This is called an ordered basis. For example, the canonical ordered basis for $\mathbb{R}^n$ is $(e_1, e_2, \ldots, e_n)$. The possibility to reorder basis vectors is not the only way in which bases are non-unique.

Bases are not unique. While there exists a unique way to express a vector in terms of any particular basis, bases themselves are far from unique. For example, both of the sets

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \text{ and } \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

are bases for $\mathbb{R}^2$. Rescaling any vector in one of these sets is already enough to show that $\mathbb{R}^2$ has infinitely many bases. But even if we require that all of the basis vectors have unit length, it turns out that there are still infinitely many bases for $\mathbb{R}^2$ (see review question 3).

To see whether a set of vectors $S = \{v_1, \ldots, v_m\}$ is a basis for $\mathbb{R}^n$, we have to check that the elements are linearly independent and that they span $\mathbb{R}^n$. From the previous discussion, we also know that m must equal n, so lets assume S has n vectors. If S is linearly independent, then there is no non-trivial solution of the equation

$$0 = x^1 v_1 + \cdots + x^n v_n .$$

Let M be a matrix whose columns are the vectors v_i and X the column vector with entries x^i. Then the above equation is equivalent to requiring that there is a unique solution to

$$MX = 0 .$$

To see if S spans $\mathbb{R}^n$, we take an arbitrary vector w and solve the linear system

$$w = x^1 v_1 + \cdots + x^n v_n$$

in the unknowns x^i. For this, we need to find a unique solution for the linear system $MX = w$.

Thus, we need to show that M^{-1} exists, so that

$$X = M^{-1}w$$

is the unique solution we desire. Then we see that S is a basis for $\mathbb{R}^n$ if and only if $\det M \neq 0$.

Theorem 11.1.1. *Let $S = \{v_1, \ldots, v_m\}$ be a collection of vectors in $\mathbb{R}^n$. Let M be the matrix whose columns are the vectors in S. Then S is a basis for V if and only if m is the dimension of V and*

$$\det M \neq 0.$$

Remark Also observe that S is a basis if and only if $\mathrm{RREF}(M) = I$.

Example 122 Let

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \text{ and } T = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}.$$

Then set $M_S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Since $\det M_S = 1 \neq 0$, then S is a basis for $\mathbb{R}^2$.

Likewise, set $M_T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Since $\det M_T = -2 \neq 0$, then T is a basis for $\mathbb{R}^2$.

11.2 Matrix of a Linear Transformation (Redux)

Not only do bases allow us to describe arbitrary vectors as column vectors, they also permit linear transformations to be expressed as matrices. This is a very powerful tool for computations, which is covered in chapter 7 and reviewed again here.

Suppose we have a linear transformation $L \colon V \to W$ and ordered input and output bases $E = (e_1, \ldots, e_n)$ and $F = (f_1, \ldots, f_m)$ for V and W respectively (of course, these need not be the standard basis–in all likelihood V is *not* $\mathbb{R}^n$). Since for each e_j, $L(e_j)$ is a vector in W, there exist unique numbers m^i_j such that

$$L(e_j) = f_1 m^1_j + \cdots + f_m m^m_j = (f_1, \ldots, f_m) \begin{pmatrix} m^1_j \\ \vdots \\ m^m_j \end{pmatrix}.$$

The number m^i_j is the ith component of $L(e_j)$ in the basis F, while the f_i are vectors (note that if α is a scalar, and v a vector, $\alpha v = v\alpha$, we have used the latter—rather uncommon—notation in the above formula). The numbers m^i_j naturally form a matrix whose jth column is the column vector displayed above. Indeed, if

$$v = e_1 v^1 + \cdots + e_n v^n \,,$$

Then

$$\begin{aligned}
L(v) &= L(v^1 e_1 + v^2 e_2 + \cdots + v^n e_n) \\
&= v^1 L(e_1) + v^2 L(e_2) + \cdots + v^n L(e_n) \;=\; \sum_{j=1}^{m} L(e_j) v^j \\
&= \sum_{j=1}^{m} (f_1 m^1_j + \cdots + f_m m^m_j) v^j \;=\; \sum_{i=1}^{n} f_i \left[\sum_{j=1}^{m} M^i_j v^j \right] \\
&= \begin{pmatrix} f_1 & f_2 & \cdots & f_m \end{pmatrix} \begin{pmatrix} m^1_1 & m^1_2 & \cdots & m^1_n \\ m^2_1 & m^2_2 & & \\ \vdots & & \ddots & \vdots \\ m^m_1 & & \cdots & m^m_n \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix}
\end{aligned}$$

In the column vector-basis notation this equality looks familiar:

$$L \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}_E = \left(\begin{pmatrix} m^1_1 & \ldots & m^1_n \\ \vdots & & \vdots \\ m^m_1 & \ldots & m^m_n \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \right)_F .$$

The array of numbers $M = (m^i_j)$ is called the matrix of L in the input and output bases E and F for V and W, respectively. This matrix will change if we change either of the bases. Also observe that the columns of M are computed by examining L acting on each basis vector in V expanded in the basis vectors of W.

Example 123 Let $L \colon P_1(t) \mapsto P_1(t)$, such that $L(a + bt) = (a + b)t$. Since $V =$

$P_1(t) = W$, let's choose the same ordered basis $B = (1-t, 1+t)$ for V and W.

$$\begin{aligned} L(1-t) &= (1-1)t = \ 0 \ = (1-t)\cdot 0 + (1+t)\cdot 0 = \big(1-t, 1+t\big)\begin{pmatrix}0\\0\end{pmatrix} \\ L(1+t) &= (1+1)t = 2t \ = (1-t)\cdot -1 + (1+t)\cdot 1 = \big(1-t, 1+t\big)\begin{pmatrix}-1\\1\end{pmatrix} \\ &\Rightarrow \ L\begin{pmatrix}a\\b\end{pmatrix}_B = \left(\begin{pmatrix}0 & -1\\0 & 1\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}\right)_B . \end{aligned}$$

When the vector space is $\mathbb{R}^n$ and the standard basis is used, the problem of finding the matrix of a linear transformation will seem almost trivial. It is worthwhile working through it once in the above language though.

Example 124 Any vector in $\mathbb{R}^n$ can be written as a linear combination of the *standard (ordered) basis* $(e_1, \ldots e_n)$. The vector e_i has a one in the ith position, and zeros everywhere else. $I.e.$

$$e_1 = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \quad e_2 = \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \ldots, \quad e_n = \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}.$$

Then to find the matrix of any linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^n$, it suffices to know what $L(e_i)$ is for every i.

For any matrix M, observe that Me_i is equal to the ith column of M. Then if the ith column of M equals $L(e_i)$ for every i, then $Mv = L(v)$ for every $v \in \mathbb{R}^n$. Then the matrix representing L in the standard basis is just the matrix whose ith column is $L(e_i)$.

For example, if

$$L\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1\\4\\7\end{pmatrix}, \quad L\begin{pmatrix}0\\1\\0\end{pmatrix} = \begin{pmatrix}2\\5\\8\end{pmatrix}, \quad L\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}3\\6\\9\end{pmatrix},$$

then the matrix of L in the standard basis is simply

$$\begin{pmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{pmatrix}.$$

Alternatively, this information would often be presented as

$$L\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}x+2y+3z\\4x+5y+6z\\7x+8y+9z\end{pmatrix}.$$

You could either rewrite this as

$$L\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix},$$

to immediately learn the matrix of L, or taking a more circuitous route:

$$\begin{aligned}L\begin{pmatrix}x\\y\\z\end{pmatrix} &= L\left[x\begin{pmatrix}1\\0\\0\end{pmatrix}+y\begin{pmatrix}0\\0\\1\end{pmatrix}+z\begin{pmatrix}0\\0\\1\end{pmatrix}\right]\\ &= x\begin{pmatrix}1\\4\\7\end{pmatrix}+y\begin{pmatrix}2\\5\\8\end{pmatrix}+z\begin{pmatrix}3\\6\\9\end{pmatrix}=\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}.\end{aligned}$$

11.3 Review Problems

Webwork:

Reading Problems	1,2
Basis checks	3,4
Computing column vectors	5,6

1. (a) Draw the collection of all unit vectors in $\mathbb{R}^2$.

   (b) Let $S_x=\left\{\begin{pmatrix}1\\0\end{pmatrix},x\right\}$, where x is a unit vector in $\mathbb{R}^2$. For which x is S_x a basis of $\mathbb{R}^2$?

   (c) Sketch all unit vectors in $\mathbb{R}^3$.

   (d) For which $x\in\mathbb{R}^3$ is $S_x=\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},x\right\}$ a basis for $\mathbb{R}^3$.

   (e) Discuss the generalization of the above to $\mathbb{R}^n$.

2. Let B^n be the vector space of column vectors with bit entries $0,1$. Write down every basis for B^1 and B^2. How many bases are there for B^3? B^4? Can you make a conjecture for the number of bases for B^n?

(Hint: You can build up a basis for B^n by choosing one vector at a time, such that the vector you choose is not in the span of the previous vectors you've chosen. How many vectors are in the span of any one vector? Any two vectors? How many vectors are in the span of any k vectors, for $k \leq n$?)



Hint



3. Suppose that V is an n-dimensional vector space.

   (a) Show that any n linearly independent vectors in V form a basis.

   (Hint: Let $\{w_1, \dots, w_m\}$ be a collection of n linearly independent vectors in V, and let $\{v_1, \dots, v_n\}$ be a basis for V. Apply the method of Lemma 11.0.2 to these two sets of vectors.)

   (b) Show that any set of n vectors in V which span V forms a basis for V.

   (Hint: Suppose that you have a set of n vectors which span V but do not form a basis. What must be true about them? How could you get a basis from this set? Use Corollary 11.0.3 to derive a contradiction.)

4. Let $S = \{v_1, \dots, v_n\}$ be a subset of a vector space V. Show that if every vector w in V can be expressed uniquely as a linear combination of vectors in S, then S is a basis of V. In other words: suppose that for every vector w in V, there is exactly one set of constants $c^1, \dots, c^n$ so that $c^1 v_1 + \cdots + c^n v_n = w$. Show that this means that the set S is linearly independent and spans V. (This is the converse to theorem 11.0.1.)

5. Vectors are objects that you can add together; show that the set of all linear transformations mapping $\mathbb{R}^3 \to \mathbb{R}$ is itself a vector space. Find a basis for this vector space. Do you think your proof could be modified to work for linear transformations $\mathbb{R}^n \to \mathbb{R}$? For $\mathbb{R}^{\mathbb{N}} \to \mathbb{R}^m$? For $\mathbb{R}^{\mathbb{R}}$?

   Hint: Represent $\mathbb{R}^3$ as column vectors, and argue that a linear transformation $T \colon \mathbb{R}^3 \to \mathbb{R}$ is just a row vector.

6. Let S_n denote the vector space of all $n \times n$ symmetric matrices;

$$S_n := \{M : \mathbb{R}^n \to \mathbb{R}^n \mid M = M^T\}.$$

Let A_n denote the vector space of all $n \times n$ anti-symmetric matrices;

$$A_n = \{M : \mathbb{R}^n \to \mathbb{R}^n \mid M = -M^T\}.$$

   (a) Find a basis for S_3.
   (b) Find a basis for A_3.
   (c) Can you find a basis for S_n? For A_n?
   Hint: Describe it in terms of combinations of the matrices F^i_j which have a 1 in the i-th row and the j-th column and 0 everywhere else. Note that $\{F^i_j \mid 1 \le i \le r, 1 \le j \le k\}$ is a basis for M^r_k.

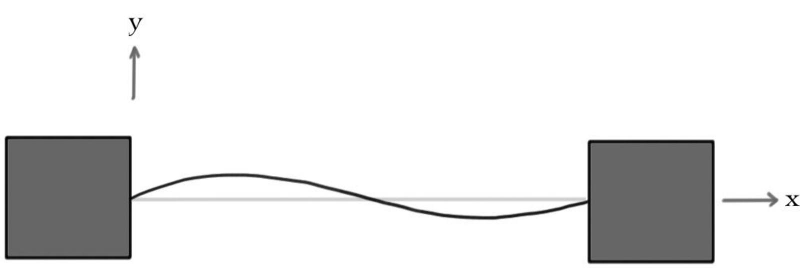
7. Give the matrix of the linear transformation L with respect to the input and output bases B and B' listed below:

   (a) $L : V \to W$ where $B = (v_1, \ldots, v_n)$ is a basis for V and $B' = (L(v_1), \ldots, L(v_n))$ is a basis for W.
   (b) $L : V \to V$ where $B = B' = (v_1, \ldots, v_n)$ and $L(v_i) = \lambda_i v_i$ for all $1 \le i \le n$.

12 Eigenvalues and Eigenvectors

In a vector space with no structure other than the vector space rules, no vector other than the zero vector is any more important than any other. Once one also has a linear transformation the situation changes dramatically. We begin with a fun example, of a type bound to reappear in your future scientific studies:

String Theory Consider a vibrating string, whose displacement at point x at time t is given by a function $y(x,t)$:



The set of all displacement functions for the string can be modeled by a vector space

$$V = \left\{ y : \mathbb{R}^2 \to \mathbb{R} \middle| \text{all partial derivatives } \frac{\partial^{k+m} y(x,t)}{\partial x^k \partial t^m} \text{ exist} \right\}.$$

The concavity and the acceleration of the string at the point (x,t) are $\frac{\partial^2 y}{\partial x^2}(x,t)$ and $\frac{\partial^2 y}{\partial t^2}(x,t)$ respectively. Since quantities must exist at each point on the string for the wave equation to make sense, we required that all partial derivatives of $y(x,t)$ exist.

Note also that the function $y(x,t) = 0$ —drawn in grey—is the only special vector in the vector space V.

We now add some extra information. The string's behavior in time and space can be modeled by a wave equation

$$\frac{\partial^2 y}{\partial t^2} = \frac{\partial^2 y}{\partial x^2},$$

which says that the acceleration of a point on the string is equal its concavity at that point. For example, if the string were made of stretched rubber, it would prefer to be in a straight line, so this equation makes good intuitive sense. Not all of the functions in V are solutions to the wave equation; not all of the functions in the vector space V describe the way a string would really vibrate. The ways a string would really vibrate are (at least approximately) solutions to the wave equation above, which can rewritten as a linear function

$$Wy = 0$$

where

$$W = \left(-\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2}\right) : V \to V\,.$$

Some examples of solutions are

$$y_1(x,t) = \sin(t)\sin(x) \qquad y_2(x,t) = 3\sin(2t)\sin(2x)$$

and

$$y_3(x,t) = \sin(t)\sin(x) + 3\sin(2t)\sin(2x)\,.$$

Since $Wy = 0$ is a homogeneous linear equation, linear combinations of solutions are solutions; in other words the kernel $\ker(w)$ is a vector space. Given the linear function W, some vectors are now more special than others.

We can use musical intuition to do more! If the ends of the string were held fixed, we suspect that it would prefer to vibrate at certain frequencies corresponding to musical notes. This is modeled by looking at solutions of the form

$$y(x,t) = \sin(\omega t)v(x)\,.$$

Here the periodic sine function accounts for the string's vibratory motion, while the function $v(x)$ gives the shape of the string at any fixed instant of time. Observe that

$$W\big(\sin(\omega t)v(x)\big) = \sin(\omega t)\big(\frac{d^2 f}{dx^2} + \omega^2 f\big)\,.$$

This suggests we introduce a new vector space

$$U = \left\{ v : \mathbb{R} \to \mathbb{R} \,\middle|\, \text{all derivatives } \frac{d^k f}{dx^k} \text{ exist} \right\},$$

as well as a new linear function

$$L := \frac{d^2}{dx^2} : U \longrightarrow U\,.$$

The number ω is called an angular frequency in many contexts, lets call its square $\lambda := -\omega^2$ to match notations we will use later (notice that for this particular problem λ must then be negative). Then, because we want $W(y) = 0$, which implies $d^2f/dx^2 = \omega^2 f$, it follows that the vector $v(x) \in U$ determining the vibrating string's shape obeys

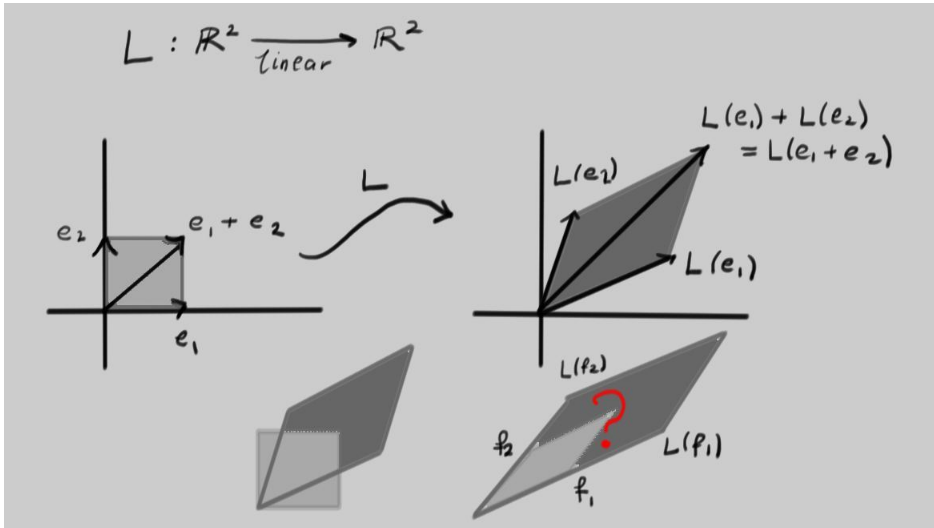
$$L(v) = \lambda v\,.$$

This is perhaps one of the most important equations in all of linear algebra! It is the eigenvalue-eigenvector equation. In this problem we have to solve it both for λ, to determine which frequencies (or musical notes) our string likes to sing, and the vector v determining the string's shape. The vector v is called an eigenvector and λ its corresponding eigenvalue. The solution sets for each λ are called V_λ. For any λ the set V_k is a vector space since elements of this set are solutions to the homogeneous equation $(L - \lambda)v = 0$.

We began this chapter by stating "In a vector space, with no other structure, no vector is more important than any other." Our aim is to show you that when a linear operator L acts on a vector space, vectors that solve the equation $L(v) = \lambda v$ play a central role.

12.1 Invariant Directions

Have a look at the linear transformation L depicted below:



It was picked at random by choosing a pair of vectors $L(e_1)$ and $L(e_2)$ as the outputs of L acting on the canonical basis vectors. Notice how the unit square with a corner at the origin is mapped to a parallelogram. The second line of the picture shows these superimposed on one another. Now look at the second picture on that line. There, two vectors f_1 and f_2 have been carefully chosen such that if the inputs into L are in the parallelogram spanned by f_1 and f_2, the outputs also form a parallelogram with edges lying along the same two directions. Clearly this is a very special situation that should correspond to interesting properties of L.

Now lets try an explicit example to see if we can achieve the last picture:

Example 126 Consider the linear transformation L such that

$$L\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}-4\\-10\end{pmatrix} \text{ and } L\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}3\\7\end{pmatrix},$$

so that the matrix of L in the standard basis is

$$\begin{pmatrix}-4 & 3\\-10 & 7\end{pmatrix}.$$

Recall that a vector is a direction and a magnitude; L applied to $\begin{pmatrix}1\\0\end{pmatrix}$ or $\begin{pmatrix}0\\1\end{pmatrix}$ changes both the direction and the magnitude of the vectors given to it.

Notice that

$$L\begin{pmatrix}3\\5\end{pmatrix} = \begin{pmatrix}-4\cdot 3+3\cdot 5\\-10\cdot 3+7\cdot 5\end{pmatrix} = \begin{pmatrix}3\\5\end{pmatrix}.$$

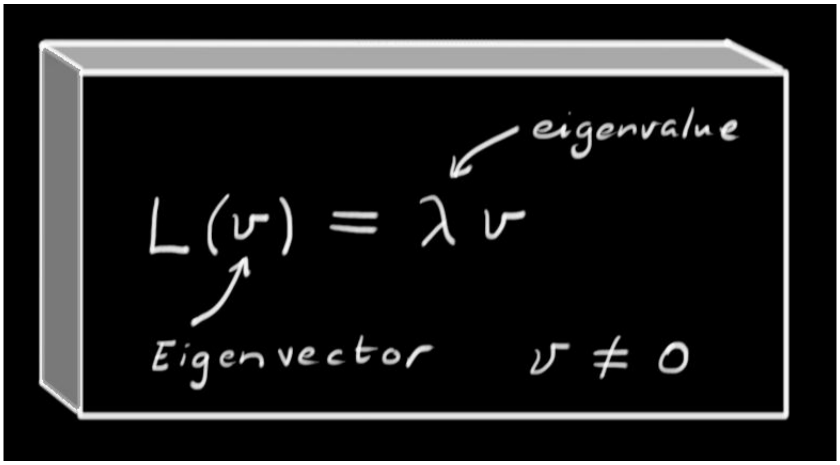


Figure 12.1: The eigenvalue–eigenvector equation is probably the most important one in linear algebra.

Then L fixes the direction (and actually also the magnitude) of the vector $v_1 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$.

Reading homework: problem 1

Now, notice that any vector with the same direction as v_1 can be written as cv_1 for some constant c. Then $L(cv_1) = cL(v_1) = cv_1$, so L fixes every vector pointing in the same direction as v_1.

Also notice that

$$L\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -4 \cdot 1 + 3 \cdot 2 \\ -10 \cdot 1 + 7 \cdot 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

so L fixes the direction of the vector $v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ but stretches v_2 by a factor of 2. Now notice that for any constant c, $L(cv_2) = cL(v_2) = 2cv_2$. Then L stretches every vector pointing in the same direction as v_2 by a factor of 2.

In short, given a linear transformation L it is sometimes possible to find a vector $v \neq 0$ and constant $\lambda \neq 0$ such that $Lv = \lambda v$. We call the direction of the vector v an *invariant direction*. In fact, any vector pointing in the same

direction also satisfies this equation because $L(cv) = cL(v) = \lambda cv$. More generally, any *non-zero* vector v that solves

$$L(v) = \lambda v$$

is called an *eigenvector* of L, and λ (which now need not be zero) is an *eigenvalue*. Since the direction is all we really care about here, then any other vector cv (so long as $c \neq 0$) is an equally good choice of eigenvector. Notice that the relation "u and v point in the same direction" is an equivalence relation.

In our example of the linear transformation L with matrix

$$\begin{pmatrix} -4 & 3 \\ -10 & 7 \end{pmatrix},$$

we have seen that L enjoys the property of having two invariant directions, represented by eigenvectors v_1 and v_2 with eigenvalues 1 and 2, respectively.

It would be very convenient if we could write any vector w as a linear combination of v_1 and v_2. Suppose $w = rv_1 + sv_2$ for some constants r and s. Then

$$L(w) = L(rv_1 + sv_2) = rL(v_1) + sL(v_2) = rv_1 + 2sv_2.$$

Now L just multiplies the number r by 1 and the number s by 2. If we could write this as a matrix, it would look like:

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix}$$

which is much slicker than the usual scenario

$$L\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}.$$

Here, s and t give the coordinates of w in terms of the vectors v_1 and v_2. In the previous example, we multiplied the vector by the matrix L and came up with a complicated expression. In these coordinates, we see that L has a very simple *diagonal matrix*, whose diagonal entries are exactly the *eigenvalues* of L.

This process is called *diagonalization*. It makes complicated linear systems much easier to analyze.

Reading homework: problem 2

Now that we've seen what eigenvalues and eigenvectors are, there are a number of questions that need to be answered.

- How do we find eigenvectors and their eigenvalues?
- How many eigenvalues and (independent) eigenvectors does a given linear transformation have?
- When can a linear transformation be diagonalized?

We will start by trying to find the eigenvectors for a linear transformation.

2×2 Example

Example 127 Let $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $L(x,y) = (2x+2y, 16x+6y)$. First, we find the matrix of L, this is quickest in the standard basis:

$$\begin{pmatrix} x \\ y \end{pmatrix} \stackrel{L}{\longmapsto} \begin{pmatrix} 2 & 2 \\ 16 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

We want to find an invariant direction $v = \begin{pmatrix} x \\ y \end{pmatrix}$ such that

$$Lv = \lambda v$$

or, in matrix notation,

$$\begin{aligned}
\begin{pmatrix} 2 & 2 \\ 16 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \lambda \begin{pmatrix} x \\ y \end{pmatrix} \\
\Leftrightarrow \qquad \begin{pmatrix} 2 & 2 \\ 16 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
\Leftrightarrow \qquad \begin{pmatrix} 2-\lambda & 2 \\ 16 & 6-\lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\end{aligned}$$

This is a homogeneous system, so it only has solutions when the matrix $\begin{pmatrix}2-\lambda & 2\\ 16 & 6-\lambda\end{pmatrix}$ is singular. In other words,

$$\begin{array}{rrcl} & \det\begin{pmatrix}2-\lambda & 2\\ 16 & 6-\lambda\end{pmatrix} & = & 0\\ \Leftrightarrow & (2-\lambda)(6-\lambda)-32 & = & 0\\ \Leftrightarrow & \lambda^2-8\lambda-20 & = & 0\\ \Leftrightarrow & (\lambda-10)(\lambda+2) & = & 0\end{array}$$

For any square $n\times n$ matrix M, the polynomial in λ given by

$$P_M(\lambda)=\det(\lambda I-M)=(-1)^n\det(M-\lambda I)$$

is called the *characteristic polynomial* of M, and its roots are the eigenvalues of M.

In this case, we see that L has two eigenvalues, $\lambda_1=10$ and $\lambda_2=-2$. To find the eigenvectors, we need to deal with these two cases separately. To do so, we solve the linear system $\begin{pmatrix}2-\lambda & 2\\ 16 & 6-\lambda\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$ with the particular eigenvalue λ plugged in to the matrix.

$\underline{\lambda=10}$: We solve the linear system

$$\begin{pmatrix}-8 & 2\\ 16 & -4\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}.$$

Both equations say that $y=4x$, so any vector $\begin{pmatrix}x\\ 4x\end{pmatrix}$ will do. Since we only need the direction of the eigenvector, we can pick a value for x. Setting $x=1$ is convenient, and gives the eigenvector $v_1=\begin{pmatrix}1\\ 4\end{pmatrix}$.

$\underline{\lambda=-2}$: We solve the linear system

$$\begin{pmatrix}4 & 2\\ 16 & 8\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}.$$

Here again both equations agree, because we chose λ to make the system singular. We see that $y=-2x$ works, so we can choose $v_2=\begin{pmatrix}1\\ -2\end{pmatrix}$.

Our process was the following:

1. Find the characteristic polynomial of the matrix M for L, given by[1] $\det(\lambda I - M)$.
2. Find the roots of the characteristic polynomial; these are the eigenvalues of L.
3. For each eigenvalue λ_i, solve the linear system $(M - \lambda_i I)v = 0$ to obtain an eigenvector v associated to λ_i.

Jordan block example

12.2 The Eigenvalue–Eigenvector Equation

In section 12, we developed the idea of eigenvalues and eigenvectors in the case of linear transformations $\mathbb{R}^2 \to \mathbb{R}^2$. In this section, we will develop the idea more generally.

Eigenvalues

Definition If $L: V \to V$ is linear and for some scalar λ and $v \neq 0_V$

$$Lv = \lambda v.$$

then λ is an **eigenvalue** of L with **eigenvector** v.

This equation says that the direction of v is invariant (unchanged) under L.

Let's try to understand this equation better in terms of matrices. Let V be a finite-dimensional vector space and let $L: V \to V$. If we have a basis for V we can represent L by a square matrix M and find eigenvalues λ and associated eigenvectors v by solving the homogeneous system

$$(M - \lambda I)v = 0.$$

This system has non-zero solutions if and only if the matrix

$$M - \lambda I$$

is singular, and so we require that

[1] To save writing many minus signs compute $\det(M - \lambda I)$; which is equivalent if you only need the roots.

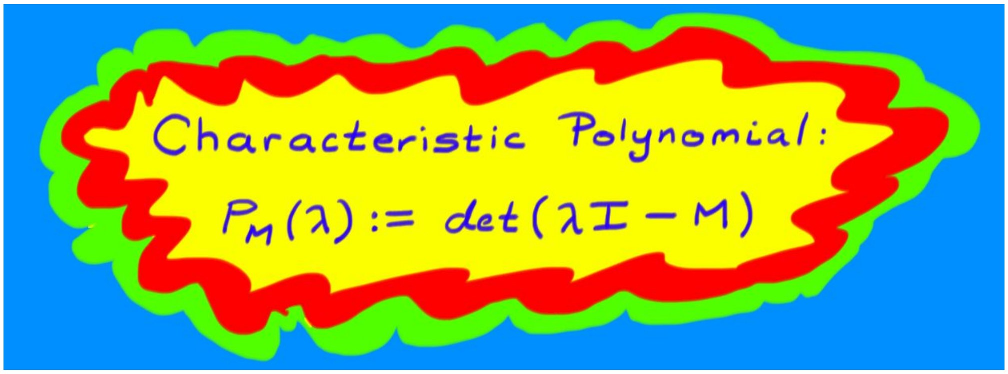


Figure 12.2: Don't forget the characteristic polynomial; you will need it to compute eigenvalues.

$$\det(\lambda I - M) = 0.$$

The left hand side of this equation is a polynomial in the variable λ called the **characteristic polynomial** $P_M(\lambda)$ of M. For an $n \times n$ matrix, the characteristic polynomial has degree n. Then

$$P_M(\lambda) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_n.$$

Notice that $P_M(0) = \det(-M) = (-1)^n \det M$.

Now recall the following.

Theorem 12.2.1. *(The Fundamental Theorem of Algebra) Any polynomial can be factored into a product of first order polynomials over $\mathbb{C}$.*

This theorem implies that there exists a collection of n complex numbers λ_i (possibly with repetition) such that

$$P_M(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) \implies P_M(\lambda_i) = 0.$$

The eigenvalues λ_i of M are exactly the roots of $P_M(\lambda)$. These eigenvalues could be real or complex or zero, and they need not all be different. The number of times that any given root λ_i appears in the collection of eigenvalues is called its *multiplicity*.

Example 128 Let L be the linear transformation $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$L\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}2x+y-z\\x+2y-z\\-x-y+2z\end{pmatrix}.$$

In the standard basis the matrix M representing L has columns Le_i for each i, so:

$$\begin{pmatrix}x\\y\\z\end{pmatrix} \stackrel{L}{\longmapsto} \begin{pmatrix}2&1&-1\\1&2&-1\\-1&-1&2\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}.$$

Then the characteristic polynomial of L is[2]

$$\begin{aligned}P_M(\lambda) &= \det\begin{pmatrix}\lambda-2&-1&1\\-1&\lambda-2&1\\1&1&\lambda-2\end{pmatrix}\\ &= (\lambda-2)[(\lambda-2)^2-1]+[-(\lambda-2)-1]+[-(\lambda-2)-1]\\ &= (\lambda-1)^2(\lambda-4)\,.\end{aligned}$$

So L has eigenvalues $\lambda_1 = 1$ (with multiplicity 2), and $\lambda_2 = 4$ (with multiplicity 1).

To find the eigenvectors associated to each eigenvalue, we solve the homogeneous system $(M - \lambda_i I)X = 0$ for each i.

$\underline{\lambda = 4}$: We set up the augmented matrix for the linear system:

$$\left(\begin{array}{rrr|r}-2&1&-1&0\\1&-2&-1&0\\-1&-1&-2&0\end{array}\right) \sim \left(\begin{array}{rrr|r}1&-2&-1&0\\0&-3&-3&0\\0&-3&-3&0\end{array}\right) \sim \left(\begin{array}{rrr|r}1&0&1&0\\0&1&1&0\\0&0&0&0\end{array}\right).$$

Any vector of the form $t\begin{pmatrix}-1\\-1\\1\end{pmatrix}$ is then an eigenvector with eigenvalue 4; thus L leaves a line through the origin invariant.

[2] It is often easier (and equivalent) to solve $\det(M - \lambda I) = 0$.

$\underline{\lambda = 1}$: Again we set up an augmented matrix and find the solution set:

$$\left(\begin{array}{rrr|r} 1 & 1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ -1 & -1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Then the solution set has two free parameters, s and t, such that $z = z =: t$, $y = y =: s$, and $x = -s + t$. Thus L leaves invariant the set:

$$\left\{ s \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \middle| s, t \in \mathbb{R} \right\}.$$

This set is a plane through the origin. So the multiplicity two eigenvalue has two independent eigenvectors, $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ that determine an invariant plane.

Example 129 Let V be the vector space of smooth (*i.e.* infinitely differentiable) functions $f \colon \mathbb{R} \to \mathbb{R}$. Then the derivative is a linear operator $\frac{d}{dx} \colon V \to V$. What are the eigenvectors of the derivative? In this case, we don't have a matrix to work with, so we have to make do.

A function f is an eigenvector of $\frac{d}{dx}$ if there exists some number λ such that

$$\frac{d}{dx} f = \lambda f .$$

An obvious candidate is the exponential function, $e^{\lambda x}$; indeed, $\frac{d}{dx} e^{\lambda x} = \lambda e^{\lambda x}$. The operator $\frac{d}{dx}$ has an eigenvector $e^{\lambda x}$ for every $\lambda \in \mathbb{R}$.

12.3 Eigenspaces

In the previous example, we found two eigenvectors

$$\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

for L, both with eigenvalue 1. Notice that

$$\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

is also an eigenvector of L with eigenvalue 1. In fact, any linear combination

$$r\begin{pmatrix}-1\\1\\0\end{pmatrix}+s\begin{pmatrix}1\\0\\1\end{pmatrix}$$

of these two eigenvectors will be another eigenvector with the same eigenvalue.

More generally, let $\{v_1, v_2, \ldots\}$ be eigenvectors of some linear transformation L with the same eigenvalue λ. A *linear combination* of the v_i is given by $c^1v_1 + c^2v_2 + \cdots$ for some constants $c^1, c^2, \ldots$. Then

$$\begin{aligned} L(c^1v_1 + c^2v_2 + \cdots) &= c^1Lv_1 + c^2Lv_2 + \cdots \text{ by linearity of } L \\ &= c^1\lambda v_1 + c^2\lambda v_2 + \cdots \text{ since } Lv_i = \lambda v_i \\ &= \lambda(c^1v_1 + c^2v_2 + \cdots). \end{aligned}$$

So every linear combination of the v_i is an eigenvector of L with the same eigenvalue λ. In simple terms, any sum of eigenvectors is again an eigenvector *if they share the same eigenvalue.*

The space of all vectors with eigenvalue λ is called an **eigenspace**. It is, in fact, a vector space contained within the larger vector space V. It contains 0_V, since $L0_V = 0_V = \lambda 0_V$, and is closed under addition and scalar multiplication by the above calculation. All other vector space properties are inherited from the fact that V itself is a vector space. In other words, the subspace theorem (9.1.1, chapter 9) ensures that $V_\lambda := \{v \in V | Lv = 0\}$ is a subspace of V.



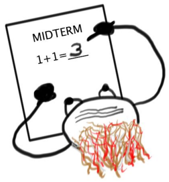
Eigenspaces





Reading homework: problem 3

You can now attempt the second sample midterm.



12.4 Review Problems

Webwork:

Reading Problems	1, 2, 3
Characteristic polynomial	4, 5, 6
Eigenvalues	7, 8
Eigenspaces	9, 10
Eigenvectors	11, 12, 13, 14
Complex eigenvalues	15

1. Try to find more solutions to the vibrating string problem $\partial^2 y/\partial t^2 = \partial^2 y/\partial x^2$ using the ansatz

$$y(x,t) = \sin(\omega t) f(x)\,.$$

What equation must $f(x)$ obey? Can you write this as an eigenvector equation? Suppose that the string has length L and $f(0) = f(L) = 0$. Can you find any solutions for $f(x)$?

2. Let $M = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$. Find all eigenvalues of M. Does M have two linearly independent eigenvectors? Is there a basis in which the matrix of M is diagonal? (*I.e.*, can M be diagonalized?)

3. Consider $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ with

$$L\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta \end{pmatrix}.$$

   (a) Write the matrix of L in the basis $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

   (b) When $\theta \neq 0$, explain how L acts on the plane. Draw a picture.

   (c) Do you expect L to have invariant directions? (Consider also special values of θ.)

   (d) Try to find real eigenvalues for L by solving the equation

$$L(v) = \lambda v.$$

   (e) Are there complex eigenvalues for L, assuming that $i = \sqrt{-1}$ exists?

4. Let L be the linear transformation $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+y \\ x+z \\ y+z \end{pmatrix}.$$

   Let e_i be the vector with a one in the ith position and zeros in all other positions.

   (a) Find Le_i for each $i = 1, 2, 3$.

   (b) Given a matrix $M = \begin{pmatrix} m^1_1 & m^1_2 & m^1_3 \\ m^2_1 & m^2_2 & m^2_3 \\ m^3_1 & m^3_2 & m^3_3 \end{pmatrix}$, what can you say about Me_i for each i?

   (c) Find a 3×3 matrix M representing L.

   (d) Find the eigenvectors and eigenvalues of M.

5. Let A be a matrix with eigenvector v with eigenvalue λ. Show that v is also an eigenvector for A^2 and find the corresponding eigenvalue. How about for A^n where $n \in \mathbb{N}$? Suppose that A is invertible. Show that v is also an eigenvector for A^{-1}.

6. A *projection* is a linear operator P such that $P^2 = P$. Let v be an eigenvector with eigenvalue λ for a projection P, what are all possible values of λ? Show that every projection P has at least one eigenvector.

   Note that every complex matrix has at least 1 eigenvector, but you need to prove the above for *any* field.

7. Explain why the characteristic polynomial of an $n \times n$ matrix has degree n. Make your explanation easy to read by starting with some simple examples, and then use properties of the determinant to give a *general* explanation.

8. Compute the characteristic polynomial $P_M(\lambda)$ of the matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

   Now, since we can evaluate polynomials on square matrices, we can plug M into its characteristic polynomial and find the *matrix* $P_M(M)$.

What do you find from this computation? Does something similar hold for 3×3 matrices? (Try assuming that the matrix of M is diagonal to answer this.)

9. *Discrete dynamical system.* Let M be the matrix given by

$$M = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}.$$

Given any vector $v(0) = \begin{pmatrix} x(0) \\ y(0) \end{pmatrix}$, we can create an infinite sequence of vectors $v(1), v(2), v(3)$, and so on using the rule:

$$v(t+1) = Mv(t) \text{ for all natural numbers } t.$$

(This is known as a *discrete dynamical system* whose *initial condition* is $v(0)$.)

(a) Find all eigenvectors and eigenvalues of M.

(b) Find all vectors $v(0)$ such that

$$v(0) = v(1) = v(2) = v(3) = \cdots$$

(Such a vector is known as a *fixed point* of the dynamical system.)

(c) Find all vectors $v(0)$ such that $v(0), v(1), v(2), v(3), \ldots$ all point in the same direction. (Any such vector describes an *invariant curve* of the dynamical system.)



Hint



13 Diagonalization

Given a linear transformation, it is highly desirable to write its matrix with respect to a basis of eigenvectors.

13.1 Diagonalizability

Suppose we are lucky, and we have $L\colon V \to V$, and the ordered basis $B = (v_1, \ldots, v_n)$ is a set of eigenvectors for L, with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then:

$$\begin{array}{rcl} L(v_1) & = & \lambda_1 v_1 \\ L(v_2) & = & \lambda_2 v_2 \\ & \vdots & \\ L(v_n) & = & \lambda_n v_n \end{array}$$

As a result, the matrix of L in the basis of eigenvectors B is diagonal:

$$L \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix}_B = \left(\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix} \right)_B ,$$

where all entries off the diagonal are zero.

Suppose that V is any n-dimensional vector space. We call a linear transformation $L: V \mapsto V$ *diagonalizable* if there exists a collection of n linearly independent eigenvectors for L. In other words, L is diagonalizable if there exists a basis for V of eigenvectors for L.

In a basis of eigenvectors, the matrix of a linear transformation is diagonal. On the other hand, if an $n \times n$ matrix is diagonal, then the standard basis vectors e_i must already be a set of n linearly independent eigenvectors. We have shown:

Theorem 13.1.1. *Given an ordered basis B for a vector space V and a linear transformation $L: V \rightarrow V$, then the matrix for L in the basis B is diagonal if and only if B consists of eigenvectors for L.*

`Non-diagonalizable example`

Reading homework: problem 1

Typically, however, we do not begin a problem with a basis of eigenvectors, but rather have to compute these. Hence we need to know how to change from one basis to another:

13.2 Change of Basis

Suppose we have two ordered bases $S = (v_1, \ldots, v_n)$ and $S' = (v'_1, \ldots, v'_n)$ for a vector space V. (Here v_i and v'_i are *vectors*, not components of vectors in a basis!) Then we may write each v'_k uniquely as

$$v'_k = \sum_i v_i p^i_k \,,$$

this is v'_k as a linear combination of the v_i. In matrix notation

$$\begin{pmatrix} v'_1, v'_2, \cdots, v'_n \end{pmatrix} = \begin{pmatrix} v_1, v_2, \cdots, v_n \end{pmatrix} \begin{pmatrix} p^1_1 & p^1_2 & \cdots & p^1_n \\ p^2_1 & p^2_2 & & \\ \vdots & & & \vdots \\ p^n_1 & & \cdots & p^n_n \end{pmatrix} .$$

Here, the p^i_k are constants, which we can regard as entries of a square matrix $P = (p^i_k)$. The matrix P must have an inverse since we can also write each v_j uniquely as a linear combination of the v'_k;

$$v_j = \sum_k v'_k q^k_j .$$

Then we can write

$$v_j = \sum_k \sum_i v_i p^i_k q^k_j .$$

But $\sum_k p^i_k q^k_j$ is the k, j entry of the product matrix PQ. Since the expression for v_j in the basis S is v_j itself, then PQ maps each v_j to itself. As a result, each v_j is an eigenvector for PQ with eigenvalue 1, so PQ is the identity, *i.e.*

$$PQ = I \Leftrightarrow Q = P^{-1} .$$

The matrix P is called a *change of basis* matrix. There is a quick and dirty trick to obtain it; look at the formula above relating the new basis vectors $v'_1, v'_2, \ldots v'_n$ to the old ones $v_1, v_2, \ldots, v_n$. In particular focus on v'_1 for which

$$v'_1 = \begin{pmatrix} v_1, v_2, \cdots, v_n \end{pmatrix} \begin{pmatrix} p^1_1 \\ p^2_1 \\ \vdots \\ p^n_1 \end{pmatrix} .$$

This says that the first column of the change of basis matrix P is really just the components of the vector v'_1 in the basis $v_1, v_2, \ldots, v_n$.

> The columns of the change of basis matrix are the components of the new basis vectors in terms of the old basis vectors.

Example 130 Suppose $S' = (v'_1, v'_2)$ is an ordered basis for a vector space V and that with respect to some other ordered basis $S = (v_1, v_2)$ for V

$$v'_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}_S \quad \text{and} \quad v'_2 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix}_S .$$

This means

$$v_1' = (v_1, v_2)\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{v_1 + v_2}{\sqrt{2}} \quad \text{and} \quad v_2' = (v_1, v_2)\begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{pmatrix} = \frac{v_1 - v_2}{\sqrt{3}}\,.$$

The change of basis matrix has as its columns just the components of v_1' and v_2';

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix}.$$

Changing basis changes the matrix of a linear transformation. However, as a map between vector spaces, **the linear transformation is the same no matter which basis we use**. Linear transformations are the actual objects of study of this book, not matrices; matrices are merely a convenient way of doing computations.

Change of Basis Example

Lets now calculate how the matrix of a linear transformation changes when changing basis. To wit, let $L\colon V \longrightarrow W$ with matrix $M = (m^i_j)$ in the ordered input and output bases $S = (v_1, \ldots, v_n)$ and $T = (w_1, \ldots, w_m)$ so

$$L(v_i) = \sum_k w_k m^k_i.$$

Now, suppose $S' = (v_1', \ldots, v_n')$ and $T' = (w_1', \ldots, w_m')$ are new ordered input and out bases with matrix $M' = ({m'}^k_i)$. Then

$$L(v_i') = \sum_k w_k {m'}^k_i\,.$$

Let $P = (p^i_j)$ be the change of basis matrix from input basis S to the basis S' and $Q = (q^j_k)$ be the change of basis matrix from output basis T to the basis T'. Then:

$$L(v_j') = L\left(\sum_i v_i p^i_j\right) = \sum_i L(v_i) p^i_j = \sum_i \sum_k w_k m^k_i p^i_j.$$

Meanwhile, we have:

$$L(v'_i) = \sum_k v_k m'^k_i = \sum_k \sum_j v_j q^j_k m^k_i .$$

Since the expression for a vector in a basis is unique, then we see that the entries of MP are the same as the entries of QM'. In other words, we see that

$$MP = QM' \qquad \text{or} \qquad M' = Q^{-1}MP.$$

Example 131 Let V be the space of polynomials in t and degree 2 or less and $L : V \to \mathbb{R}^2$ where

$$L(1) = \begin{pmatrix}1\\2\end{pmatrix} \quad L(t) = \begin{pmatrix}2\\1\end{pmatrix} , \quad L(t^2) = \begin{pmatrix}3\\3\end{pmatrix} .$$

From this information we can immediately read off the matrix M of L in the bases $S = (1, t, t^2)$ and $T = (e_1, e_2)$, the standard basis for $\mathbb{R}^2$, because

$$\begin{aligned}\big(L(1), L(t), L(t^2)\big) &= (e_1 + 2e_2, 2e_1 + e_2, 3e_1 + 3e_2)\\ &= (e_1, e_2)\begin{pmatrix}1 & 2 & 3\\ 2 & 1 & 3\end{pmatrix} \Rightarrow M = \begin{pmatrix}1 & 2 & 3\\ 2 & 1 & 3\end{pmatrix} .\end{aligned}$$

Now suppose we are more interested in the bases

$$S' = (1 + t, t + t^2, 1 + t^2) , \quad T' = \left(\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}\right) =: (w'_1, w'_2) .$$

To compute the new matrix M' of L we could simply calculate what L does the the new input basis vectors in terms of the new output basis vectors:

$$\begin{aligned}\big(L(1 + t), L(t + t^2), L(1 + t^2)\big) &= \left(\begin{pmatrix}1\\2\end{pmatrix} + \begin{pmatrix}2\\1\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix} + \begin{pmatrix}3\\3\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix} + \begin{pmatrix}3\\3\end{pmatrix}\right)\\ &= (w'_1 + w'_2, w'_1 + 2w'_2, 2w'_1 + w'_2)\\ &= (w'_1, w'_2)\begin{pmatrix}1 & 1 & 2\\ 1 & 2 & 1\end{pmatrix} \Rightarrow M' = \begin{pmatrix}1 & 1 & 2\\ 1 & 2 & 1\end{pmatrix} .\end{aligned}$$

Alternatively we could calculate the change of basis matrices P and Q by noting that

$$(1 + t, t + t^2, 1 + t^2) = (1, t, t^2)\begin{pmatrix}1 & 0 & 1\\ 1 & 1 & 0\\ 0 & 1 & 1\end{pmatrix} \Rightarrow P = \begin{pmatrix}1 & 0 & 1\\ 1 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}$$

and

$$(w'_1, w'_2) = (e_1 + 2e_2, 2e_1 + e_2) = (e_1, e_1)\begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix} \Rightarrow Q = \begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}.$$

Hence

$$M' = Q^{-1}MP = -\frac{1}{3}\begin{pmatrix}1 & -2\\ -2 & 1\end{pmatrix}\begin{pmatrix}1 & 2 & 3\\ 2 & 1 & 3\end{pmatrix}\begin{pmatrix}1 & 0 & 1\\ 1 & 1 & 0\\ 0 & 1 & 1\end{pmatrix} = \begin{pmatrix}1 & 1 & 2\\ 1 & 2 & 1\end{pmatrix}.$$

Notice that the change of basis matrices P and Q are both square and invertible. Also, since we really wanted Q^{-1}, it is more efficient to try and write (e_1, e_2) in terms of (w'_1, w'_2) which would yield directly Q^{-1}. Alternatively, one can check that $MP = QM'$.

13.3 Changing to a Basis of Eigenvectors

If we are changing to a basis of eigenvectors, then there are various simplifications:

- Since $L : V \to V$, most likely you already know the matrix M of L using the same input basis as output basis $S = (u_1, \ldots, u_n)$ (say).
- In the new basis of eigenvectors $S'(v_1, \ldots, v_n)$, the matrix D of L is diagonal because $Lv_i = \lambda_i v_i$ and so

$$\big(L(v_1), L(v_2), \ldots, L(v_n)\big) = (v_1, v_2, \ldots, v_n)\begin{pmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & & 0\\ \vdots & & \ddots & \vdots\\ 0 & 0 & \cdots & \lambda_n\end{pmatrix}.$$

- If P is the change of basis matrix from S to S', the diagonal matrix of eigenvalues D and the original matrix are related by

$$\boxed{D = P^{-1}MP}$$

This motivates the following definition:

Definition A matrix M is **diagonalizable** if there exists an invertible matrix P and a diagonal matrix D such that

$$D = P^{-1}MP.$$

We can summarize as follows.

- Change of basis rearranges the components of a vector by the change of basis matrix P, to give components in the new basis.
- To get the matrix of a linear transformation in the new basis, we *conjugate* the matrix of L by the change of basis matrix: $M \mapsto P^{-1}MP$.

If for two matrices N and M there exists a matrix P such that $M = P^{-1}NP$, then we say that M and N are **similar**. Then the above discussion shows that diagonalizable matrices are similar to diagonal matrices.

Corollary 13.3.1. *A square matrix M is diagonalizable if and only if there exists a basis of eigenvectors for M. Moreover, these eigenvectors are the columns of a change of basis matrix P which diagonalizes M.*

Reading homework: problem 2

Example 132 Let's try to diagonalize the matrix

$$M = \begin{pmatrix} -14 & -28 & -44 \\ -7 & -14 & -23 \\ 9 & 18 & 29 \end{pmatrix}.$$

The eigenvalues of M are determined by

$$\det(M - \lambda I) = -\lambda^3 + \lambda^2 + 2\lambda = 0.$$

So the eigenvalues of M are $-1, 0,$ and 2, and associated eigenvectors turn out to be

$$v_1 = \begin{pmatrix} -8 \\ -1 \\ 3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \text{ and } \quad v_3 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}.$$

In order for M to be diagonalizable, we need the vectors v_1, v_2, v_3 to be linearly independent. Notice that the matrix

$$P = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} = \begin{pmatrix} -8 & -2 & -1 \\ -1 & 1 & -1 \\ 3 & 0 & 1 \end{pmatrix}$$

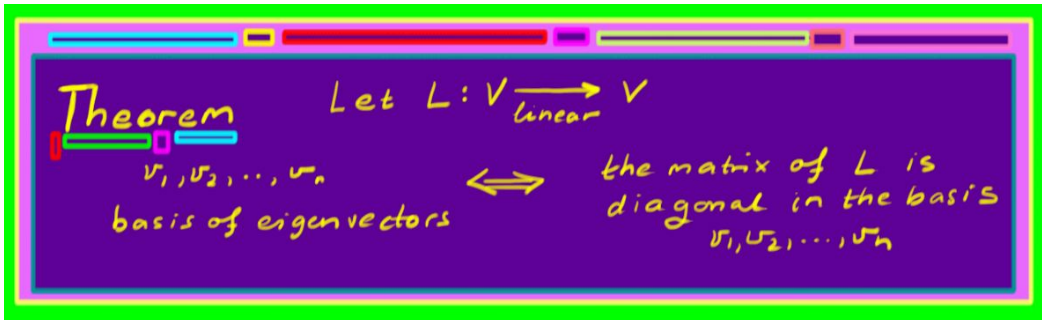


Figure 13.1: This theorem answers the question: "What is diagonalization?"

is invertible because its determinant is -1. Therefore, the eigenvectors of M form a basis of $\mathbb{R}$, and so M is diagonalizable. Moreover, because the columns of P are the components of eigenvectors,

$$MP = \begin{pmatrix} Mv_1 & Mv_2 & Mv_3 \end{pmatrix} = \begin{pmatrix} -1.v_1 & 0.v_2 & 2.v_3 \end{pmatrix} = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Hence, the matrix P of eigenvectors is a change of basis matrix that diagonalizes M;

$$P^{-1}MP = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$



2×2 Example



13.4 Review Problems

Webwork:

Reading Problems	1, 2
No real eigenvalues	3
Diagonalization	4, 5, 6, 7

1. Let $P_n(t)$ be the vector space of polynomials of degree n or less, and $\frac{d}{dt}: P_n(t) \to P_n(t)$ be the derivative operator. Find the matrix of $\frac{d}{dt}$ in the ordered bases $E = (1, t, \ldots, t^n)$ for the domain and $F =$

$(t^n, \ldots, , t, 1)$ for the codomain. Determine if this derivative operator is diagonalizable.

Recall from Chapter 6 that the derivative operator is linear .

2. When writing a matrix for a linear transformation, we have seen that the choice of basis matters. In fact, even the order of the basis matters!

   (a) Write all possible reorderings of the standard basis (e_1, e_2, e_3) for $\mathbb{R}^3$.

   (b) Write each change of basis matrix between the standard basis and each of its reorderings. Make as many observations as you can about these matrices. what are their entries? Do you notice anything about how many of each type of entry appears in each row and column? What are their determinants? (Note: These matrices are known as *permutation matrices*.)

   (c) Given $L : \mathbb{R}^3 \to \mathbb{R}^3$ is linear and

$$L \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2y - z \\ 3x \\ 2z + x + y \end{pmatrix}$$

   write the matrix M for L in the standard basis, and two reorderings of the standard basis. How are these matrices related?

3. Let

$$X = \{\heartsuit, \clubsuit, \spadesuit\}, \quad Y = \{*, \star\}.$$

   Write down two different ordered bases, S, S' and T, T' respectively, for each of the vector spaces $\mathbb{R}^X$ and $\mathbb{R}^Y$. Find the change of basis matrices P and Q that map these bases to one another. Now consider the map

$$\ell : Y \to X,$$

   where $\ell(*) = \heartsuit$ and $\ell(\star) = \spadesuit$. Show that ℓ can be used to define a linear transformation $L : \mathbb{R}^X \to \mathbb{R}^Y$. Compute the matrices M and M' of L in the bases S, T and then S', T'. Use your change of basis matrices P and Q to check that $M' = Q^{-1}MP$.

4. Recall that $\operatorname{tr} MN = \operatorname{tr} NM$. Use this fact to show that the trace of a square matrix M does not depend on the basis you used to compute M.

5. When is the 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ diagonalizable? Include examples in your answer.

6. Show that similarity of matrices is an *equivalence relation*. (The definition of an equivalence relation is given in the background WeBWorK set.)

7. *Jordan form*

   - Can the matrix $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ be diagonalized? Either diagonalize it or explain why this is impossible.
   - Can the matrix $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$ be diagonalized? Either diagonalize it or explain why this is impossible.
   - Can the $n \times n$ matrix $\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & & 0 & \cdots & 0 & \lambda \end{pmatrix}$ be diagonalized? Either diagonalize it or explain why this is impossible.

     Note: It turns out that every matrix is similar to a block matrix whose diagonal blocks look like diagonal matrices or the ones above and whose off-diagonal blocks are all zero. This is called the *Jordan form* of the matrix and a (maximal) block that looks like

     $$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & & 0 \\ \vdots & & \ddots & \ddots & \\ & & & \lambda & 1 \\ 0 & 0 & & 0 & \lambda \end{pmatrix}$$

     is called a *Jordan n-cell* or a *Jordan block* where n is the size of the block.

8. Let A and B be commuting matrices (*i.e.*, $AB = BA$) and suppose that A has an eigenvector v with eigenvalue λ.

(a) Show that Bv is also an eigenvector of A with eigenvalue λ.

(b) Additionally suppose that A is diagonalizable with distinct eigenvalues. What is the dimension of each eigenspace of A?

(c) Show that v is also an eigenvector of B.

(d) Explain why this shows that A and B can be *simultaneously diagonalized* (*i.e.* there is an ordered basis in which both their matrices are diagonal).

14 Orthonormal Bases and Complements

You may have noticed that we have only rarely used the dot product. That is because many of the results we have obtained do not require a preferred notion of lengths of vectors. Once a dot or inner product is available, lengths of and angles between vectors can be measured–very powerful machinery and results are available in this case.

14.1 Properties of the Standard Basis

The standard notion of the length of a vector $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ is

$$||x|| = \sqrt{x \cdot x} = \sqrt{(x_1)^2 + (x_2)^2 + \cdots (x_n)^2}\,.$$

The canonical/standard basis in $\mathbb{R}^n$

$$e_1 = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \quad e_2 = \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \quad \ldots, \quad e_n = \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix},$$

has many useful properties with respect to the dot product and lengths.

- Each of the standard basis vectors has unit length;

$$\|e_i\| = \sqrt{e_i \cdot e_i} = \sqrt{e_i^T e_i} = 1\,.$$

- The standard basis vectors are *orthogonal* (in other words, at right angles or perpendicular);

$$e_i \cdot e_j = e_i^T e_j = 0 \text{ when } i \neq j$$

This is summarized by

$$e_i^T e_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases},$$

where δ_{ij} is the *Kronecker delta*. Notice that the Kronecker delta gives the entries of the identity matrix.

Given column vectors v and w, we have seen that the dot product $v \cdot w$ is the same as the matrix multiplication $v^T w$. This is an *inner product* on $\mathbb{R}^n$. We can also form the *outer product* vw^T, which gives a square matrix. The outer product on the standard basis vectors is interesting. Set

$$\begin{aligned}
\Pi_1 &= e_1 e_1^T \\
&= \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \\
&\vdots \\
\Pi_n &= e_n e_n^T \\
&= \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & \cdots & 1 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}
\end{aligned}$$

In short, Π_i is the diagonal square matrix with a 1 in the ith diagonal position and zeros everywhere else[1].

Notice that $\Pi_i \Pi_j = e_i e_i^T e_j e_j^T = e_i \delta_{ij} e_j^T$. Then:

$$\Pi_i \Pi_j = \begin{cases} \Pi_i & i = j \\ 0 & i \neq j \end{cases} .$$

Moreover, for a diagonal matrix D with diagonal entries $\lambda_1, \ldots, \lambda_n$, we can write

$$D = \lambda_1 \Pi_1 + \cdots + \lambda_n \Pi_n .$$

14.2 Orthogonal and Orthonormal Bases

There are many other bases that behave in the same way as the standard basis. As such, we will study:

- *Orthogonal bases* $\{v_1, \ldots, v_n\}$:

  $$v_i \cdot v_j = 0 \text{ if } i \neq j .$$

  In other words, all vectors in the basis are perpendicular.

- *Orthonormal bases* $\{u_1, \ldots, u_n\}$:

  $$u_i \cdot u_j = \delta_{ij} .$$

  In addition to being orthogonal, each vector has unit length.

Suppose $T = \{u_1, \ldots, u_n\}$ is an orthonormal basis for $\mathbb{R}^n$. Because T is a basis, we can write any vector v uniquely as a linear combination of the vectors in T;

$$v = c^1 u_1 + \cdots c^n u_n .$$

Since T is orthonormal, there is a very easy way to find the coefficients of this linear combination. By taking the dot product of v with any of the vectors

[1] This is reminiscent of an older notation, where vectors are written in juxtaposition. This is called a "dyadic tensor", and is still used in some applications.

in T, we get

$$
\begin{aligned}
v \cdot u_i &= c^1 u_1 \cdot u_i + \cdots + c^i u_i \cdot u_i + \cdots + c^n u_n \cdot u_i \\
&= c^1 \cdot 0 + \cdots + c^i \cdot 1 + \cdots + c^n \cdot 0 \\
&= c^i , \\
\Rightarrow c^i &= v \cdot u_i \\
\Rightarrow v &= (v \cdot u_1) u_1 + \cdots + (v \cdot u_n) u_n \\
&= \sum_i (v \cdot u_i) u_i \, .
\end{aligned}
$$

This proves the following theorem.

Theorem 14.2.1. *For an orthonormal basis $\{u_1, \ldots, u_n\}$, any vector v can be expressed as*

$$v = \sum_i (v \cdot u_i) u_i .$$

Reading homework: problem 1

All orthonormal bases for $\mathbb{R}^2$

14.2.1 Orthonormal Bases and Dot Products

To calculate lengths of, and angles between vectors in $\mathbb{R}^n$ we most commonly use the dot product:

$$\begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} \cdot \begin{pmatrix} w^1 \\ \vdots \\ w^n \end{pmatrix} := v^1 w^1 + \cdots + v^n w^n \, .$$

When dealing with more general vector spaces the dot product makes no sense, and one must instead choose an appropriate inner product. By "appropriate", we mean an inner product well-suited to the problem one is trying to solve. If the vector space V under study has an orthonormal basis $O = (u_1, \ldots, u_n)$ meaning

$$\langle u_i, u_j \rangle = \delta_{ij} \, ,$$

where $\langle \cdot, \cdot \rangle$ is the inner product, you might ask whether this can be related to a dot product? The answer to this question is yes and rather easy to understand:

Given an orthonormal basis, the information of two vectors v and v' in V can be encoded in column vectors

$$\begin{aligned} v &= \langle v, u_1\rangle u_1 + \cdots + \langle v, u_n\rangle u_n = (u_1, \ldots, u_n) \begin{pmatrix} \langle v, u_1\rangle \\ \vdots \\ \langle v, u_n\rangle \end{pmatrix} = \begin{pmatrix} \langle v, u_1\rangle \\ \vdots \\ \langle v, u_n\rangle \end{pmatrix}_O , \\ v' &= \langle v', u_1\rangle u_1 + \cdots + \langle v', u_n\rangle u_n = (u_1, \ldots, u_n) \begin{pmatrix} \langle v', u_1\rangle \\ \vdots \\ \langle v', u_n\rangle \end{pmatrix} = \begin{pmatrix} \langle v', u_1\rangle \\ \vdots \\ \langle v', u_n\rangle \end{pmatrix}_O . \end{aligned}$$

The dot product of these two column vectors is

$$\begin{pmatrix} \langle v, u_1\rangle \\ \vdots \\ \langle v, u_n\rangle \end{pmatrix} \cdot \begin{pmatrix} \langle v', u_1\rangle \\ \vdots \\ \langle v', u_n\rangle \end{pmatrix} = \langle v, u_1\rangle\langle v', u_1\rangle + \cdots + \langle v, u_n\rangle\langle v,' u_n\rangle .$$

This agrees exactly with the inner product of v and v' because

$$\begin{aligned} \langle v, v'\rangle &= \Big\langle \langle v, u_1\rangle u_1 + \cdots + \langle v, u_n\rangle u_n, \langle v', u_1\rangle u_1 + \cdots + \langle v', u_n\rangle u_n \Big\rangle \\ &= \langle v, u_1\rangle\langle v', u_1\rangle\langle u_1, u_1\rangle + \langle v, u_2\rangle\langle v', u_1\rangle\langle u_2, u_1\rangle + \cdots \\ &\quad \cdots + \langle v, u_{n-1}\rangle\langle v', u_n\rangle\langle u_{n-1}, u_n\rangle + \langle v, u_n\rangle\langle v', u_n\rangle\langle u_n, u_n\rangle \\ &= \langle v, u_1\rangle\langle v', u_1\rangle + \cdots + \langle v, u_n\rangle\langle v', u_n\rangle . \end{aligned}$$

The above computation looks a little daunting, but only the linearity property of inner products and the fact that $\langle u_i, u_j\rangle$ can equal either zero or one was used. Because inner products become dot products once one uses an orthonormal basis, we will quite often use the dot product notation in situations where one really should write an inner product. Conversely, dot product computations can always be rewritten in terms of an inner product, if needed.

Example 133 Consider the space of polynomials given by $V = \mathrm{span}\{1, x\}$ with inner product $\langle p, p'\rangle = \int_0^1 p(x)p'(x)dx$. An obvious basis to use is $B = (1, x)$ but it is not hard to check that this is not orthonormal, instead we take

$$O = \Big(1, 2\sqrt{3}\,\big(x - \frac{1}{2}\big)\Big) .$$

This is an orthonormal basis since,for example:

$$\Big\langle 2\sqrt{3}(x-\frac{1}{2}),1\Big\rangle = 2\sqrt{3}\int_0^1 (x-\frac{1}{2})dx = 0\,,$$

and

$$\Big\langle x-\frac{1}{2},x-\frac{1}{2}\Big\rangle = \int_0^1 (x-\frac{1}{2})^2 dx = \frac{1}{12} = \Big(\frac{1}{2\sqrt{3}}\Big)^2\,.$$

An arbitrary vector $v = a + bx$ is given in terms of the orthonormal basis O by

$$v = (a+\frac{b}{2}).1 + b(x-\frac{1}{2}) = \Big(1, 2\sqrt{3}\,(x-\frac{1}{2})\Big)\begin{pmatrix} a+\frac{b}{2} \\ \frac{b}{2\sqrt{3}} \end{pmatrix} = \begin{pmatrix} a+\frac{b}{2} \\ \frac{b}{2\sqrt{3}} \end{pmatrix}_O\,.$$

Hence we can predict the inner product of $a + bx$ and $a' + b'x$ using the dot product:

$$\begin{pmatrix} a+\frac{b}{2} \\ \frac{b}{2\sqrt{3}} \end{pmatrix} \cdot \begin{pmatrix} a'+\frac{b'}{2} \\ \frac{b'}{2\sqrt{3}} \end{pmatrix} = \Big(a+\frac{b}{2}\Big)\Big(a'+\frac{b'}{2}\Big) + \frac{bb'}{12} = aa' + \frac{1}{2}\,(ab'+a'b) + \frac{1}{3}\,bb'\,.$$

Indeed

$$\langle a+bx, a'+b'x\rangle = \int_0^1 (a+bx)(a'+b'x)dx = aa' + \frac{1}{2}\,(ab'+a'b) + \frac{1}{3}\,bb'\,.$$

14.3 Relating Orthonormal Bases

Suppose $T = \{u_1, \dots, u_n\}$ and $R = \{w_1, \dots, w_n\}$ are two orthonormal bases for $\mathbb{R}^n$. Then

$$\begin{aligned} w_1 &= (w_1 \cdot u_1)u_1 + \cdots + (w_1 \cdot u_n)u_n \\ &\vdots \\ w_n &= (w_n \cdot u_1)u_1 + \cdots + (w_n \cdot u_n)u_n \\ \Rightarrow w_i &= \sum_j u_j(u_j \cdot w_i) \end{aligned}$$

Thus the matrix for the change of basis from T to R is given by

$$P = (p^j_i) = (u_j \cdot w_i).$$

We would like to calculate the product PP^T. For that, we first develop a dirty trick for products of dot products:

$$(u \cdot v)(w \cdot z) = (u^T v)(w^T z) = u^T (vw^T) z\,.$$

The object vw^T is the square matrix made from the outer product of v and w. Now we are ready to compute the components of the matrix product PP^T.

$$\begin{aligned}
\sum_i (u_j \cdot w_i)(w_i \cdot u_k) &= \sum_i (u_j^T w_i)(w_i^T u_k) \\
&= u_j^T \left[\sum_i (w_i w_i^T) \right] u_k \\
&\overset{(*)}{=} u_j^T I_n u_k \\
&= u_j^T u_k = \delta_{jk}.
\end{aligned}$$

The equality $(*)$ is explained below. Assuming $(*)$ holds, we have shown that $PP^T = I_n$, which implies that

$$P^T = P^{-1}.$$

The equality in the line $(*)$ says that $\sum_i w_i w_i^T = I_n$. To see this, we examine $\left(\sum_i w_i w_i^T\right) v$ for an arbitrary vector v. We can find constants c^j such that $v = \sum_j c^j w_j$, so that

$$\begin{aligned}
\left(\sum_i w_i w_i^T \right) v &= \left(\sum_i w_i w_i^T \right) \left(\sum_j c^j w_j \right) \\
&= \sum_j c^j \sum_i w_i w_i^T w_j \\
&= \sum_j c^j \sum_i w_i \delta_{ij} \\
&= \sum_j c^j w_j \text{ since all terms with } i \neq j \text{ vanish} \\
&= v.
\end{aligned}$$

Thus, as a linear transformation, $\sum_i w_i w_i^T = I_n$ fixes every vector, and thus must be the identity I_n.

Definition A matrix P is **orthogonal** if $P^{-1}=P^T$.

Then to summarize,

Theorem 14.3.1. *A change of basis matrix P relating two orthonormal bases is an orthogonal matrix. I.e.,*

$$\boxed{P^{-1}=P^T\,.}$$

Reading homework: problem 2

Example 134 Consider $\mathbb{R}^3$ with the ordered orthonormal basis

$$S=(u_1,u_2,u_3)=\left(\begin{pmatrix}\frac{2}{\sqrt{6}}\\ \frac{1}{\sqrt{6}}\\ \frac{-1}{\sqrt{6}}\end{pmatrix},\begin{pmatrix}0\\ \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix},\begin{pmatrix}\frac{1}{\sqrt{3}}\\ \frac{-1}{\sqrt{3}}\\ \frac{1}{\sqrt{3}}\end{pmatrix}\right).$$

Let E be the standard basis (e_1,e_2,e_3). Since we are changing from the standard basis to a new basis, then the columns of the change of basis matrix are exactly the new basis vectors. Then the change of basis matrix from E to S is given by

$$\begin{aligned}P=(P_i^j)=(e_j\cdot u_i) &= \begin{pmatrix}e_1\cdot u_1 & e_1\cdot u_2 & e_1\cdot u_3\\ e_2\cdot u_1 & e_2\cdot u_2 & e_2\cdot u_3\\ e_3\cdot u_1 & e_3\cdot u_2 & e_3\cdot u_3\end{pmatrix}\\ =\begin{pmatrix}u_1 & u_2 & u_3\end{pmatrix} &= \begin{pmatrix}\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{3}}\\ \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}}\end{pmatrix}.\end{aligned}$$

From our theorem, we observe that

$$\begin{aligned}P^{-1}=P^T &= \begin{pmatrix}u_1^T\\ u_2^T\\ u_3^T\end{pmatrix}\\ &= \begin{pmatrix}\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{6}}\\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\end{pmatrix}.\end{aligned}$$

We can check that $P^TP = I$ by a lengthy computation, or more simply, notice that

$$\begin{aligned}(P^TP) &= \begin{pmatrix} u_1^T \\ u_2^T \\ u_3^T \end{pmatrix} \begin{pmatrix} u_1 & u_2 & u_3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.\end{aligned}$$

Above we are using orthonormality of the u_i and the fact that matrix multiplication amounts to taking dot products between rows and columns. It is also very important to realize that the columns of an *orthogonal* matrix are made from an *orthonormal* set of vectors.

Orthonormal Change of Basis and Diagonal Matrices. Suppose D is a diagonal matrix and we are able to use an orthogonal matrix P to change to a new basis. Then the matrix M of D in the new basis is:

$$M = PDP^{-1} = PDP^T.$$

Now we calculate the transpose of M.

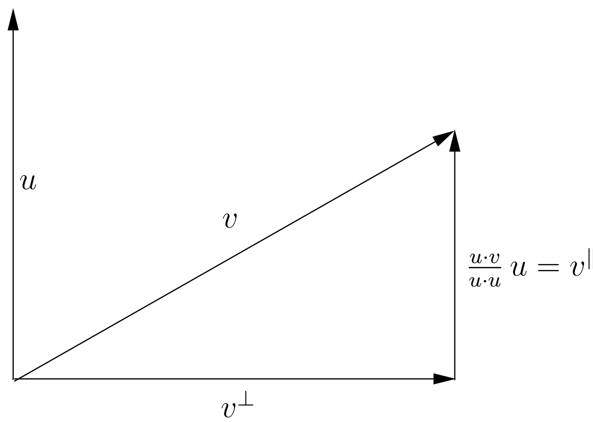
$$\begin{aligned} M^T &= (PDP^T)^T \\ &= (P^T)^T D^T P^T \\ &= PDP^T \\ &= M \end{aligned}$$

The matrix $M = PDP^T$ is symmetric!

14.4 Gram-Schmidt & Orthogonal Complements

Given a vector v and some other vector u not in span $\{v\}$ we can construct the new vector

$$\boxed{v^\perp := v - \frac{u \cdot v}{u \cdot u}\, u\,.}$$



This new vector $v^{\perp}$ is orthogonal to u because

$$u \cdot v^{\perp} = u \cdot v - \frac{u \cdot v}{u \cdot u} u \cdot u = 0.$$

Hence, $\{u, v^{\perp}\}$ is an orthogonal basis for span$\{u, v\}$. When v is not parallel to u, $v^{\perp} \neq 0$, and normalizing these vectors we obtain $\left\{\frac{u}{|u|}, \frac{v^{\perp}}{|v^{\perp}|}\right\}$, an orthonormal basis for the vector space span $\{u, v\}$.

Sometimes we write $v = v^{\perp} + v^{\|}$ where:

$$\begin{aligned} v^{\perp} &= v - \frac{u \cdot v}{u \cdot u} u \\ v^{\|} &= \frac{u \cdot v}{u \cdot u} u. \end{aligned}$$

This is called an **orthogonal decomposition** because we have decomposed v into a sum of orthogonal vectors. This decomposition depends on u; if we change the direction of u we change $v^{\perp}$ and $v^{\|}$.

If u, v are linearly independent vectors in $\mathbb{R}^3$, then the set $\{u, v^{\perp}, u \times v^{\perp}\}$ would be an orthogonal basis for $\mathbb{R}^3$. This set could then be normalized by dividing each vector by its length to obtain an orthonormal basis.

However, it often occurs that we are interested in vector spaces with dimension greater than 3, and must resort to craftier means than cross products to obtain an orthogonal basis[2].

[2] Actually, given a set T of $(n-1)$ independent vectors in n-space, one can define an analogue of the cross product that will produce a vector orthogonal to the span of T, using a method exactly analogous to the usual computation for calculating the cross product of two vectors in $\mathbb{R}^3$. This only gets us the "*last*" orthogonal vector, though; the Gram–Schmidt process described in this section gives a way to get a full orthogonal basis.

Given a third vector w, we should first check that w does not lie in the span$\{u, v\}$, *i.e.*, check that u, v and w are linearly independent. If it does not, we then can define

$$w^{\perp} := w - \frac{u \cdot w}{u \cdot u}\, u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}}\, v^{\perp}.$$

We can check that $u \cdot w^{\perp}$ and $v^{\perp} \cdot w^{\perp}$ are both zero:

$$\begin{aligned}
u \cdot w^{\perp} &= u \cdot \left(w - \frac{u \cdot w}{u \cdot u}\, u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}}\, v^{\perp} \right) \\
&= u \cdot w - \frac{u \cdot w}{u \cdot u} u \cdot u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}} u \cdot v^{\perp} \\
&= u \cdot w - u \cdot w - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}} u \cdot v^{\perp} \; = \; 0
\end{aligned}$$

since u is orthogonal to $v^{\perp}$, and

$$\begin{aligned}
v^{\perp} \cdot w^{\perp} &= v^{\perp} \cdot \left(w - \frac{u \cdot w}{u \cdot u}\, u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}}\, v^{\perp} \right) \\
&= v^{\perp} \cdot w - \frac{u \cdot w}{u \cdot u} v^{\perp} \cdot u - \frac{v^{\perp} \cdot w}{v^{\perp} \cdot v^{\perp}} v^{\perp} \cdot v^{\perp} \\
&= v^{\perp} \cdot w - \frac{u \cdot w}{u \cdot u} v^{\perp} \cdot u - v^{\perp} \cdot w \; = \; 0
\end{aligned}$$

because u is orthogonal to $v^{\perp}$. Since $w^{\perp}$ is orthogonal to both u and $v^{\perp}$, we have that $\{u, v^{\perp}, w^{\perp}\}$ is an orthogonal basis for span$\{u, v, w\}$.

14.4.1 The Gram-Schmidt Procedure

In fact, given an ordered set $(v_1, v_2, \ldots)$ of linearly independent vectors, we can define an orthogonal basis for $\operatorname{span}\{v_1, v_2, \ldots\}$ consisting of the vectors

$$\begin{aligned}
v_1^\perp &:= v_1 \\
v_2^\perp &:= v_2 - \frac{v_1^\perp \cdot v_2}{v_1^\perp \cdot v_1^\perp}\, v_1^\perp \\
v_3^\perp &:= v_3 - \frac{v_1^\perp \cdot v_3}{v_1^\perp \cdot v_1^\perp}\, v_1^\perp - \frac{v_2^\perp \cdot v_3}{v_2^\perp \cdot v_2^\perp}\, v_2^\perp \\
&\vdots \\
v_i^\perp &:= v_i - \frac{v_1^\perp \cdot v_i}{v_1^\perp \cdot v_1^\perp}\, v_1^\perp - \frac{v_2^\perp \cdot v_i}{v_2^\perp \cdot v_2^\perp}\, v_2^\perp - \cdots - \frac{v_{i-1}^\perp \cdot v_i}{v_{i-1}^\perp \cdot v_{i-1}^\perp}\, v_{i-1}^\perp \\
&\vdots
\end{aligned}$$

Notice that each $v_i^\perp$ here depends on $v_j^\perp$ for every $j < i$. This allows us to inductively/algorithmically build up a linearly independent, orthogonal set of vectors $\{v_1^\perp, v_2^\perp, \ldots\}$ such that $\operatorname{span}\{v_1^\perp, v_2^\perp, \ldots\} = \operatorname{span}\{v_1, v_2, \ldots\}$. That is, an orthogonal basis for the latter vector space.

Note that the set of vectors you start out with needs to be ordered to uniquely specify the algorithm; changing the order of the vectors will give a different orthogonal basis. You might need to be the one to put an order on the initial set of vectors.

This algorithm is called the **Gram–Schmidt orthogonalization procedure**–Gram worked at a Danish insurance company over one hundred years ago, Schmidt was a student of Hilbert (the famous German mathmatician).

Example 135 We'll obtain an orthogonal basis for $\mathbb{R}^3$ by appling Gram-Schmidt to the linearly independent set $\left\{ \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}3\\1\\1\end{pmatrix} \right\}$.

Because he Gram-Schmidt algorithm uses the first vector from the ordered set the largest number of times, we will choose the vector with the most zeros to be the first in hopes of simplifying computations; we choose to order the set as

$$(v_1, v_2, v_3) := \left(\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}3\\1\\1\end{pmatrix} \right).$$

First, we set $v_1^\perp := v_1$. Then

$$v_2^\perp := \begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{2}{2}\begin{pmatrix}1\\1\\0\end{pmatrix} = \begin{pmatrix}0\\0\\1\end{pmatrix}$$

$$v_3^\perp := \begin{pmatrix}3\\1\\1\end{pmatrix} - \frac{4}{2}\begin{pmatrix}1\\1\\0\end{pmatrix} - \frac{1}{1}\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}1\\-1\\0\end{pmatrix}.$$

Then the set

$$\left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix}\right\}$$

is an orthogonal basis for $\mathbb{R}^3$. To obtain an ortho*normal* basis we simply divide each of these vectors by its length, yielding

$$\left\{\begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}\frac{1}{\sqrt{2}}\\\frac{-1}{\sqrt{2}}\\0\end{pmatrix}\right\}.$$

A 4×4 Gram--Schmidt Example

14.5 QR Decomposition

In Chapter 7, Section 7.7 teaches you how to solve linear systems by decomposing a matrix M into a product of lower and upper triangular matrices

$$M = LU\,.$$

The Gram–Schmidt procedure suggests another matrix decomposition,

$$M = QR\,,$$

where Q is an orthogonal matrix and R is an upper triangular matrix. So-called QR-decompositions are useful for solving linear systems, eigenvalue problems and least squares approximations. You can easily get the idea behind the QR decomposition by working through a simple example.

Example 136 Find the QR decomposition of

$$M = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 3 & -2 \\ 0 & 1 & -2 \end{pmatrix}.$$

What we will do is to think of the columns of M as three 3-vectors and use Gram–Schmidt to build an orthonormal basis from these that will become the columns of the orthogonal matrix Q. We will use the matrix R to record the steps of the Gram–Schmidt procedure in such a way that the product QR equals M.

To begin with we write

$$M = \begin{pmatrix} 2 & -\frac{7}{5} & 1 \\ 1 & \frac{14}{5} & -2 \\ 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{5} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In the first matrix the first two columns are orthogonal because we simply replaced the second column of M by the vector that the Gram–Schmidt procedure produces from the first two columns of M, namely

$$\begin{pmatrix} -\frac{7}{5} \\ \frac{14}{5} \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix} - \frac{1}{5} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

The matrix on the right is almost the identity matrix, save the $+\frac{1}{5}$ in the second entry of the first row, whose effect upon multiplying the two matrices precisely undoes what we we did to the second column of the first matrix.

For the third column of M we use Gram–Schmidt to deduce the third orthogonal vector

$$\begin{pmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ -\frac{7}{6} \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ -2 \end{pmatrix} - 0 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} - \frac{-9}{\frac{54}{5}} \begin{pmatrix} -\frac{7}{5} \\ \frac{14}{5} \\ 1 \end{pmatrix},$$

and therefore, using exactly the same procedure write

$$M = \begin{pmatrix} 2 & -\frac{7}{5} & -\frac{1}{6} \\ 1 & \frac{14}{5} & \frac{1}{3} \\ 0 & 1 & -\frac{7}{6} \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{5} & 0 \\ 0 & 1 & -\frac{5}{6} \\ 0 & 0 & 1 \end{pmatrix}.$$

This is not quite the answer because the first matrix is now made of mutually orthogonal column vectors, but a *bona fide* orthogonal matrix is comprised of *orthonormal*

vectors. To achieve that we divide each column of the first matrix by its length and multiply the corresponding row of the second matrix by the same amount:

$$M = \begin{pmatrix} \frac{2\sqrt{5}}{5} & -\frac{7\sqrt{30}}{90} & -\frac{\sqrt{6}}{18} \\ \frac{\sqrt{5}}{5} & \frac{7\sqrt{30}}{45} & \frac{\sqrt{6}}{9} \\ 0 & \frac{\sqrt{30}}{18} & -\frac{7\sqrt{6}}{18} \end{pmatrix} \begin{pmatrix} \sqrt{5} & \frac{\sqrt{5}}{5} & 0 \\ 0 & \frac{3\sqrt{30}}{5} & -\frac{\sqrt{30}}{2} \\ 0 & 0 & \frac{\sqrt{6}}{2} \end{pmatrix} = QR\,.$$

Geometrically what has happened here is easy to see. We started with three vectors given by the columns of M and rotated them such that the first lies along the x-axis, the second in the xy-plane and the third in some other generic direction (here it happens to be in the yz-plane).

A nice check of the above result is to verify that entry (i, j) of the matrix R equals the dot product of the i-th column of Q with the j-th column of M. (Some people memorize this fact and use it as a recipe for computing QR decompositions.) *A good test of your own understanding is to work out why this is true!*

Another QR decomposition example

14.6 Orthogonal Complements

Let U and V be subspaces of a vector space W. In Review Exercise 2, Chapter 9, you are asked to show that $U \cap V$ is a subspace of W, and that $U \cup V$ is *not* a subspace. However, $\operatorname{span}(U \cup V)$ *is* certainly a subspace, since the span of *any* subset of a vector space is a subspace. Notice that all elements of $\operatorname{span}(U \cup V)$ take the form $u + v$ with $u \in U$ and $v \in V$. We call the subspace

$$U + V := \operatorname{span}(U \cup V) = \{u + v \mid u \in U, v \in V\}$$

the *sum* of U and V. Here, we are not adding vectors, but vector spaces to produce a new vector space.

Example 137

$$\operatorname{span}\left\{\begin{pmatrix}1\\1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\\0\end{pmatrix}\right\} + \operatorname{span}\left\{\begin{pmatrix}0\\1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\1\end{pmatrix}\right\} = \operatorname{span}\left\{\begin{pmatrix}1\\1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\1\end{pmatrix}\right\}.$$

Notice that the addends have elements in common; $\begin{pmatrix}0\\1\\1\\0\end{pmatrix}$ is in both addends. Even though both of the addends are 2-dimensional their sum is not 4-dimensional.

In the special case that U and V do not have any non-zero vectors in common, their sum is a vector space with dimension $\dim U + \dim V$.

Definition If U and V are subspaces of a vector space W such that $U \cap V = \{0_W\}$ then the vector space

$$U \oplus V := \operatorname{span}(U \cup V) = \{u + v \mid u \in U, v \in V\}$$

is the **direct sum** of U and V.

Remark

- When $U \cap V = \{0_W\}$, $U + V = U \oplus V$.
- When $U \cap V \neq \{0_W\}$, $U + V \neq U \oplus V$.

This distinction is important because the direct sum has a very nice property:

Theorem 14.6.1. *If $w \in U \oplus V$ then there is only one way to write w as the sum of a vector in U and a vector in V.*

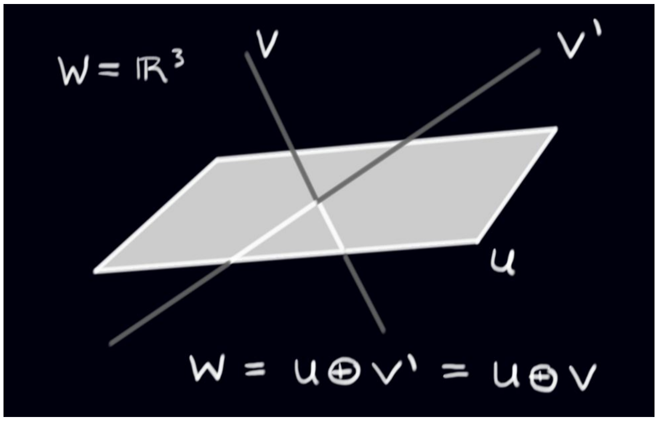
Proof. Suppose that $u + v = u' + v'$, with $u, u' \in U$, and $v, v' \in V$. Then we could express $0 = (u - u') + (v - v')$. Then $(u - u') = -(v - v')$. Since U and V are subspaces, we have $(u - u') \in U$ and $-(v - v') \in V$. But since these elements are equal, we also have $(u - u') \in V$. Since $U \cap V = \{0\}$, then $(u - u') = 0$. Similarly, $(v - v') = 0$. Therefore $u = u'$ and $v = v'$, proving the theorem. □

Reading homework: problem 3

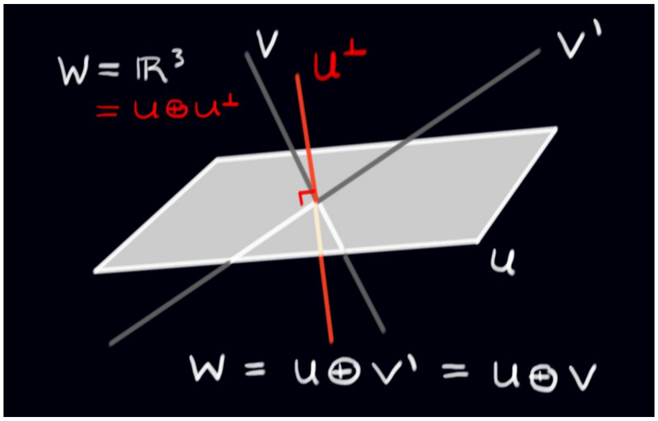
Here is a sophisticated algebra question:

> Given a subspace U in W, what are the solutions to
>
> $$U \oplus V = W.$$
>
> That is, how can we write W as the direct sum of U and *something*?

There is not a unique answer to this question as can be seen from the following picture of subspaces in $W = \mathbb{R}^3$.



However, using the inner product, there is a natural candidate $U^\perp$ for this second subspace as shown below.



Definition If U is a subspace of the vector space W then the vector space

$$U^\perp := \Big\{ w \in W \Big| w \cdot u = 0 \text{ for all } u \in U \Big\}$$

is the **orthogonal complement** of U in W.

Remark The symbols "$U^\perp$" are often read as "U-perp". This is the set of all vectors in W orthogonal to *every* vector in U. Notice also that in the above definition we have implicitly assumed that the inner product is the dot product. For a general inner product, the above definition would read $U^\perp := \Big\{ w \in W \Big| \langle w, u \rangle = 0 \text{ for all } u \in U \Big\}$.

Possibly by now you are feeling overwhelmed, it may help to watch this quick overview video.

Overview

Example 138 Consider any plane P through the origin in $\mathbb{R}^3$. Then P is a subspace, and $P^\perp$ is the line through the origin orthogonal to P. For example, if P is the xy-plane, then

$$\mathbb{R}^3 = P \oplus P^\perp = \{(x, y, 0) | x, y \in \mathbb{R}\} \oplus \{(0, 0, z) | z \in \mathbb{R}\}.$$

Theorem 14.6.2. *Let U be a subspace of a finite-dimensional vector space W. Then the set $U^\perp$ is a subspace of W, and $W = U \oplus U^\perp$.*

Proof. First, to see that $U^\perp$ is a subspace, we only need to check closure, which requires a simple check: Suppose $v, w \in U^\perp$, then we know

$$v \cdot u = 0 = w \cdot u \quad (\forall u \in U)\,.$$

Hence

$$\Rightarrow u \cdot (\alpha v + \beta w) = \alpha u \cdot v + \beta u \cdot w = 0 \quad (\forall u \in U)\,,$$

and so $\alpha v + \beta w \in U^\perp$.

Next, to form a direct sum between U and $U^\perp$ we need to show that $U \cap U^\perp = \{0\}$. This holds because if $u \in U$ and $u \in U^\perp$ it follows that

$$u \cdot u = 0 \Leftrightarrow u = 0.$$

Finally, we show that any vector $w \in W$ is in $U \oplus U^\perp$. (This is where we use the assumption that W is finite-dimensional.) Let $e_1, \ldots, e_n$ be an orthonormal basis for U. Set:

$$\begin{array}{rcl} u & = & (w \cdot e_1)e_1 + \cdots + (w \cdot e_n)e_n \in U\,, \\ u^\perp & = & w - u\,. \end{array}$$

It is easy to check that $u^\perp \in U^\perp$ (see the Gram-Schmidt procedure). Then $w = u + u^\perp$, so $w \in U \oplus U^\perp$, and we are done. $\square$

Reading homework: problem 4

Example 139 Consider any line L through the origin in $\mathbb{R}^4$. Then L is a subspace, and $L^\perp$ is a 3-dimensional subspace orthogonal to L. For example, let

$$L = \operatorname{span}\left\{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}\right\}$$

be a line in $\mathbb{R}^4$. Then

$$\begin{aligned}
L^\perp &= \left\{\begin{pmatrix}x\\y\\z\\w\end{pmatrix}\in\mathbb{R}^4 \;\middle|\; (x,y,z,w)\cdot(1,1,1,1)=0\right\}\\
&= \left\{\begin{pmatrix}x\\y\\z\\w\end{pmatrix}\in\mathbb{R}^4 \;\middle|\; x+y+z+w=0\right\}.
\end{aligned}$$

Using the Gram-Schmidt procedure one may find an orthogonal basis for $L^\perp$. The set

$$\left\{\begin{pmatrix}1\\-1\\0\\0\end{pmatrix},\begin{pmatrix}1\\0\\-1\\0\end{pmatrix},\begin{pmatrix}1\\0\\0\\-1\end{pmatrix}\right\}$$

forms a basis for $L^\perp$ so, first, we order the basis as

$$(v_1, v_2, v_2) = \left(\begin{pmatrix}1\\-1\\0\\0\end{pmatrix},\begin{pmatrix}1\\0\\-1\\0\end{pmatrix},\begin{pmatrix}1\\0\\0\\-1\end{pmatrix}\right).$$

Next, we set $v_1^\perp = v_1$. Then

$$\begin{aligned}
v_2^\perp &= \begin{pmatrix}1\\0\\-1\\0\end{pmatrix} - \frac{1}{2}\begin{pmatrix}1\\-1\\0\\0\end{pmatrix} = \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\-1\\0\end{pmatrix},\\
v_3^\perp &= \begin{pmatrix}1\\0\\0\\-1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}1\\-1\\0\\0\end{pmatrix} - \frac{1/2}{3/2}\begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\\-1\\0\end{pmatrix} = \begin{pmatrix}\frac{1}{3}\\\frac{1}{3}\\\frac{1}{3}\\-1\end{pmatrix}.
\end{aligned}$$

So the set
$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \\ -1 \end{pmatrix} \right\}$$
is an orthogonal basis for $L^{\perp}$. Dividing each basis vector by its length yields
$$\left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\sqrt{3}}{6} \\ \frac{\sqrt{3}}{6} \\ \frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{2} \end{pmatrix} \right\},$$
and orthonormal basis for $L^{\perp}$. Moreover, we have
$$\mathbb{R}^4 = L \oplus L^{\perp} = \left\{ \begin{pmatrix} c \\ c \\ c \\ c \end{pmatrix} \middle| c \in \mathbb{R} \right\} \oplus \left\{ \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \in \mathbb{R}^4 \middle| x + y + z + w = 0 \right\},$$
a decomposition of $\mathbb{R}^4$ into a line and its three dimensional orthogonal complement.

Notice that for any subspace U, the subspace $(U^{\perp})^{\perp}$ is just U again. As such, $\perp$ is an *involution* on the set of subspaces of a vector space. (An involution is any mathematical operation which performed twice does nothing.)

14.7 Review Problems

Webwork:	Reading Problems	1, 2, 3, 4
	Gram–Schmidt	5
	Orthogonal eigenbasis	6, 7
	Orthogonal complement	8

1. Let $D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

   (a) Write D in terms of the vectors e_1 and e_2, and their transposes.

   (b) Suppose $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible. Show that D is similar to
   $$M = \frac{1}{ad - bc} \begin{pmatrix} \lambda_1 ad - \lambda_2 bc & -(\lambda_1 - \lambda_2)ab \\ (\lambda_1 - \lambda_2)cd & -\lambda_1 bc + \lambda_2 ad \end{pmatrix}.$$

(c) Suppose the vectors $\begin{pmatrix}a, b\end{pmatrix}$ and $\begin{pmatrix}c, d\end{pmatrix}$ are orthogonal. What can you say about M in this case? (Hint: think about what M^T is equal to.)

2. Suppose $S = \{v_1, \ldots, v_n\}$ is an *orthogonal* (not orthonormal) basis for $\mathbb{R}^n$. Then we can write any vector v as $v = \sum_i c^i v_i$ for some constants c^i. Find a formula for the constants c^i in terms of v and the vectors in S.



Hint


3. Let u, v be linearly independent vectors in $\mathbb{R}^3$, and $P = \text{span}\{u, v\}$ be the plane spanned by u and v.

   (a) Is the vector $v^\perp := v - \frac{u \cdot v}{u \cdot u} u$ in the plane P?

   (b) What is the (cosine of the) angle between $v^\perp$ and u?

   (c) How can you find a third vector perpendicular to both u and $v^\perp$?

   (d) Construct an orthonormal basis for $\mathbb{R}^3$ from u and v.

   (e) Test your abstract formulæ starting with

$$u = \begin{pmatrix}1, 2, 0\end{pmatrix} \text{ and } v = \begin{pmatrix}0, 1, 1\end{pmatrix}.$$



Hint


4. Find an orthonormal basis for $\mathbb{R}^4$ which includes $(1, 1, 1, 1)$ using the following procedure:

   (a) Pick a vector perpendicular to the vector

$$v_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$$

from the solution set of the matrix equation

$$v_1^T x = 0\,.$$

Pick the vector v_2 obtained from the standard Gaussian elimination procedure which is the coefficient of x_2.

(b) Pick a vector perpendicular to both v_1 and v_2 from the solutions set of the matrix equation

$$\begin{pmatrix} v_1^T \\ v_2^T \end{pmatrix} x = 0\,.$$

Pick the vector v_3 obtained from the standard Gaussian elimination procedure with x_3 as the coefficient.

(c) Pick a vector perpendicular to v_1, v_2, and v_3 from the solution set of the matrix equation

$$\begin{pmatrix} v_1^T \\ v_2^T \\ v_3^T \end{pmatrix} x = 0\,.$$

Pick the vector v_4 obtained from the standard Gaussian elimination procedure with x_3 as the coefficient.

(d) Normalize the four vectors obtained above.

5. Use the inner product

$$f \cdot g := \int_0^1 f(x)g(x)dx$$

on the vector space $V = \operatorname{span}\{1, x, x^2, x^3\}$ to perform the Gram-Schmidt procedure on the set of vectors $\{1, x, x^2, x^3\}$.

6. Use the inner product

$$f \cdot g := \int_0^{2\pi} f(x)g(x)dx$$

on the vector space $V = \operatorname{span}\{\sin(x), \sin(2x), \sin(3x)\}$ to perform the Gram-Schmidt procedure on the set of vectors $\{\sin(x), \sin(2x), \sin(3x)\}$. Try to build an orthonormal basis for the vector space

$$\operatorname{span}\{\sin(nx) \mid n \in \mathbb{N}\}\,.$$

7. (a) Show that if Q is an orthogonal $n \times n$ matrix, then

$$u \cdot v = (Qu) \cdot (Qv)\,,$$

for any $u, v \in \mathbb{R}^n$. That is, Q preserves the inner product.

(b) Does Q preserve the outer product?

(c) If the set of vectors $\{u_1, \ldots, u_n\}$ is orthonormal and $\{\lambda_1, \cdots, \lambda_n\}$ is a set of numbers, then what are the eigenvalues and eigenvectors of the matrix $M = \sum_{i=1}^{n} \lambda_i u_i u_i^T$?

(d) How would the eigenvectors and eigenvalues of this matrix change if we replaced $\{u_1, \ldots, u_n\}$ by $\{Qu_1, \ldots, Qu_n\}$?

8. Carefully write out the Gram-Schmidt procedure for the set of vectors

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \right\}.$$

Is it possible to rescale the second vector obtained in the procedure to a vector with integer components?

9. (a) Suppose u and v are linearly independent. Show that u and $v^{\perp}$ are also linearly independent. Explain why $\{u, v^{\perp}\}$ is a basis for span$\{u, v\}$.



Hint



(b) Repeat the previous problem, but with three independent vectors u, v, w where $v^{\perp}$ and $w^{\perp}$ are as defined by the Gram-Schmidt procedure.

10. Find the QR factorization of

$$M = \begin{pmatrix} 1 & 0 & 2 \\ -1 & 2 & 0 \\ -1 & -2 & 2 \end{pmatrix}.$$

11. Given any three vectors u, v, w, when do $v^{\perp}$ or $w^{\perp}$ of the Gram–Schmidt procedure vanish?

12. For U a subspace of W, use the subspace theorem to check that $U^{\perp}$ is a subspace of W.

13. Let S_n and A_n define the space of $n \times n$ symmetric and anti-symmetric matrices, respectively. These are subspaces of the vector space M_n^n of all $n \times n$ matrices. What is $\dim M_n^n$, $\dim S_n$, and $\dim A_n$? Show that $M_n^n = S_n + A_n$. Define an inner product on square matrices

$$M \cdot N = \mathrm{tr} MN\,.$$

Is $A_n^{\perp} = S_n$? Is $M_n^n = S_n \oplus A_n$?

14. The vector space $V = \mathrm{span}\{\sin(t), \sin(2t), \sin(3t), \sin(3t)\}$ has an inner product:

$$f \cdot g := \int_0^{2\pi} f(t)g(t)dt\,.$$

Find the orthogonal compliment to $U = \mathrm{span}\{\sin(t) + \sin(2t)\}$ in V. Express $\sin(t) - \sin(2t)$ as the sum of vectors from U and $U^{\perp}$.

15 Diagonalizing Symmetric Matrices

Symmetric matrices have many applications. For example, if we consider the shortest distance between pairs of important cities, we might get a table like the following.

	Davis	Seattle	San Francisco
Davis	0	2000	80
Seattle	2000	0	2010
San Francisco	80	2010	0

Encoded as a matrix, we obtain

$$M = \begin{pmatrix} 0 & 2000 & 80 \\ 2000 & 0 & 2010 \\ 80 & 2010 & 0 \end{pmatrix} = M^T .$$

Definition A matrix M is **symmetric** if $M^T = M$.

One very nice property of symmetric matrices is that they always have real eigenvalues. Review exercise 1 guides you through the general proof, but below is an example for 2×2 matrices.

Example 140 For a general symmetric 2×2 matrix, we have:

$$
\begin{aligned}
P_\lambda \begin{pmatrix} a & b \\ b & d \end{pmatrix} &= \det \begin{pmatrix} \lambda - a & -b \\ -b & \lambda - d \end{pmatrix} \\
&= (\lambda - a)(\lambda - d) - b^2 \\
&= \lambda^2 - (a+d)\lambda - b^2 + ad \\
\Rightarrow \lambda &= \frac{a+d}{2} \pm \sqrt{b^2 + \left(\frac{a-d}{2}\right)^2}.
\end{aligned}
$$

Notice that the discriminant $4b^2 + (a-d)^2$ is always positive, so that the eigenvalues must be real.

Now, suppose a symmetric matrix M has two distinct eigenvalues $\lambda \neq \mu$ and eigenvectors x and y;

$$Mx = \lambda x, \qquad My = \mu y.$$

Consider the dot product $x \cdot y = x^T y = y^T x$ and calculate

$$
\begin{aligned}
x^T M y &= x^T \mu y = \mu x \cdot y, \text{ and} \\
x^T M y &= (y^T M x)^T \text{ (by transposing a } 1 \times 1 \text{ matrix)} \\
&= (y^T \lambda x)^T \\
&= (\lambda x \cdot y)^T \\
&= \lambda x \cdot y.
\end{aligned}
$$

Subtracting these two results tells us that:

$$0 = x^T M y - x^T M y = (\mu - \lambda)\, x \cdot y.$$

Since μ and λ were assumed to be distinct eigenvalues, $\lambda - \mu$ is non-zero, and so $x \cdot y = 0$. We have proved the following theorem.

Theorem 15.0.1. *Eigenvectors of a symmetric matrix with distinct eigenvalues are orthogonal.*

Reading homework: problem 1

Example 141 The matrix $M = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ has eigenvalues determined by

$$\det(M - \lambda I) = (2 - \lambda)^2 - 1 = 0.$$

So the eigenvalues of M are 3 and 1, and the associated eigenvectors turn out to be $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. It is easily seen that these eigenvectors are orthogonal;

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0.$$

In chapter 14 we saw that the matrix P built from any orthonormal basis $(v_1, \ldots, v_n)$ for $\mathbb{R}^n$ as its columns,

$$P = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} ,$$

was an orthogonal matrix. This means that

$$P^{-1} = P^T, \text{ or } PP^T = I = P^T P.$$

Moreover, given any (unit) vector x_1, one can always find vectors $x_2, \ldots, x_n$ such that $(x_1, \ldots, x_n)$ is an orthonormal basis. (Such a basis can be obtained using the Gram-Schmidt procedure.)

Now suppose M is a symmetric $n \times n$ matrix and λ_1 is an eigenvalue with eigenvector x_1 (this is always the case because every matrix has at least one eigenvalue–see Review Problem 3). Let P be the square matrix of orthonormal column vectors

$$P = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} ,$$

While x_1 is an eigenvector for M, the others are not necessarily eigenvectors for M. Then

$$MP = \begin{pmatrix} \lambda_1 x_1 & Mx_2 & \cdots & Mx_n \end{pmatrix} .$$

But P is an orthogonal matrix, so $P^{-1} = P^T$. Then:

$$\begin{aligned}
P^{-1} = P^T &= \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} \\
\Rightarrow P^T M P &= \begin{pmatrix} x_1^T \lambda_1 x_1 & * & \cdots & * \\ x_2^T \lambda_1 x_1 & * & \cdots & * \\ \vdots & & & \vdots \\ x_n^T \lambda_1 x_1 & * & \cdots & * \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & * & & \vdots \\ 0 & * & \cdots & * \end{pmatrix} \\
&= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \hat{M} & \\ 0 & & & \end{pmatrix} .
\end{aligned}$$

The last equality follows since $P^T M P$ is symmetric. The asterisks in the matrix are where "stuff" happens; this extra information is denoted by $\hat{M}$ in the final expression. We know nothing about $\hat{M}$ except that it is an $(n-1) \times (n-1)$ matrix and that it is symmetric. But then, by finding an (unit) eigenvector for $\hat{M}$, we could repeat this procedure successively. The end result would be a diagonal matrix with eigenvalues of M on the diagonal. Again, we have proved a theorem:

Theorem 15.0.2. *Every symmetric matrix is similar to a diagonal matrix of its eigenvalues. In other words,*

$$M = M^T \Leftrightarrow M = PDP^T$$

where P is an orthogonal matrix and D is a diagonal matrix whose entries are the eigenvalues of M.

Reading homework: problem 2

To diagonalize a real symmetric matrix, begin by building an orthogonal matrix from an orthonormal basis of eigenvectors, as in the example below.

Example 142 The symmetric matrix

$$M = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

has eigenvalues 3 and 1 with eigenvectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ respectively. After normalizing these eigenvectors, we build the orthogonal matrix:

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}.$$

Notice that $P^T P = I$. Then:

$$MP = \begin{pmatrix} \frac{3}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

In short, $MP = PD$, so $D = P^T MP$. Then D is the diagonalized form of M and P the associated change-of-basis matrix from the standard basis to the basis of eigenvectors.



3 × 3 Example



15.1 Review Problems

Webwork:

Reading Problems	1, 2,
Diagonalizing a symmetric matrix	3, 4

1. (On Reality of Eigenvalues)

   (a) Suppose $z = x + iy$ where $x, y \in \mathbb{R}, i = \sqrt{-1}$, and $\overline{z} = x - iy$. Compute $z\overline{z}$ and $\overline{z}z$ in terms of x and y. What kind of numbers are $z\overline{z}$ and $\overline{z}z$? (The complex number $\overline{z}$ is called the *complex conjugate* of z).

   (b) Suppose that $\lambda = x + iy$ is a complex number with $x, y \in \mathbb{R}$, and that $\lambda = \overline{\lambda}$. Does this determine the value of x or y? What kind of number must λ be?

(c) Let $x = \begin{pmatrix} z^1 \\ \vdots \\ z^n \end{pmatrix} \in \mathbb{C}^n$. Let $x^\dagger = \begin{pmatrix} \overline{z^1} & \cdots & \overline{z^n} \end{pmatrix} \in \mathbb{C}^n$ (a $1 \times n$ complex matrix or a row vector). Compute $x^\dagger x$. Using the result of part 1a, what can you say about the number $x^\dagger x$? (*E.g.,* is it real, imaginary, positive, negative, etc.)

(d) Suppose $M = M^T$ is an $n \times n$ symmetric matrix with real entries. Let λ be an eigenvalue of M with eigenvector x, so $Mx = \lambda x$. Compute:

$$\frac{x^\dagger M x}{x^\dagger x}$$

(e) Suppose Λ is a 1×1 matrix. What is Λ^T?

(f) What is the size of the matrix $x^\dagger M x$?

(g) For any matrix (or vector) N, we can compute $\overline{N}$ by applying complex conjugation to each entry of N. Compute $\overline{(x^\dagger)^T}$. Then compute $\overline{(x^\dagger M x)^T}$. Note that for matrices $\overline{AB + C} = \overline{A}\,\overline{B} + \overline{C}$.

(h) Show that $\lambda = \overline{\lambda}$. Using the result of a previous part of this problem, what does this say about λ?


Hint


2. Let

$$x_1 = \begin{pmatrix} a \\ b \\ c \end{pmatrix},$$

where $a^2 + b^2 + c^2 = 1$. Find vectors x_2 and x_3 such that $\{x_1, x_2, x_3\}$ is an orthonormal basis for $\mathbb{R}^3$. What can you say about the matrix P whose columns are the vectors x_1, x_2 and x_3 that you found?

3. Let $V \ni v \neq 0$ be a vector space, $\dim V = n$ and $L : V \xrightarrow{\text{linear}} V$.

(a) Explain why the list of vectors $(v, Lv, L^2 v, \ldots, L^n v)$ is linearly dependent.

(b) Explain why there exist scalars α_i not all zero such that

$$\alpha_0 v + \alpha_1 L v + \alpha_2 L^2 v + \cdots + \alpha_n L^n v = 0\,.$$

(c) Let m be the largest integer such that $\alpha_m \neq 0$ and

$$p(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_m z^n\,.$$

Explain why the polynomial $p(z)$ can be written as

$$p(z) = \alpha_m (z - \lambda_1)(z - \lambda_2) \ldots (z - \lambda_m)\,.$$

[Note that some of the roots λ_i could be complex.]

(d) Why does the following equation hold

$$(L - \lambda_1)(L - \lambda_2) \ldots (L - \lambda_m) v = 0\,?$$

(e) Explain why one of the numbers λ_i ($1 \leq i \leq m$) must be an eigenvalue of L.

4. (Dimensions of Eigenspaces)

(a) Let

$$A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & -2 & 2 \end{pmatrix}\,.$$

Find all eigenvalues of A.

(b) Find a basis for each eigenspace of A. What is the sum of the dimensions of the eigenspaces of A?

(c) Based on your answer to the previous part, guess a formula for the sum of the dimensions of the eigenspaces of a real $n \times n$ symmetric matrix. Explain why your formula must work for any real $n \times n$ symmetric matrix.

5. If M is not square then it can not be symmetric. However, MM^T and $M^T M$ are symmetric, and therefore diagonalizable.

(a) Is it the case that all of the eigenvalues of MM^T must also be eigenvalues of $M^T M$?

(b) Given an eigenvector of MM^T how can you obtain an eigenvector of M^TM?

(c) Let

$$M = \begin{pmatrix} 1 & 2 \\ 3 & 3 \\ 2 & 1 \end{pmatrix}.$$

Compute an orthonormal basis of eigenvectors for both MM^T and M^TM. If any of the eigenvalues for these two matrices agree, choose an order for them and use it to help order your orthonormal bases. Finally, change the input and output bases for the matrix M to these ordered orthonormal bases. Comment on what you find. (*Hint:* The result is called *the Singular Value Decomposition Theorem*.)

16 Kernel, Range, Nullity, Rank

Given a linear transformation

$$L: V \to W\,,$$

we often want to know if it has an inverse, *i.e.*, if there exists a linear transformation

$$M: W \to V$$

such that for any vector $v \in V$, we have

$$MLv = v\,,$$

and for any vector $w \in W$, we have

$$LMw = w\,.$$

A linear transformation is a special kind of function from one vector space to another. So before we discuss which linear transformations have inverses, let us first discuss inverses of arbitrary functions. When we later specialize to linear transformations, we'll also find some nice ways of creating subspaces.

Let $f: S \to T$ be a function from a set S to a set T. Recall that S is called the *domain* of f, T is called the *codomain* or *target* of f. We now formally introduce a term that should be familar to you from many previous courses.

16.1 Range

Definition The **range** of a function $f : S \to T$ is the set

$$\text{ran}(f) := \{f(s)|s \in S\} \subset T\,.$$

It is the subset of the codomain consisting of elements to which the function f maps, *i.e.*, the things in T which you can get to by starting in S and applying f.

The range of a matrix is very easy to find; the range of a matrix is the span of its columns. Thus, calculation of the range of a matrix is very easy until the last step: simplification. One aught to end by the calculation by writing the vector space as the span of a linearly independent set.

Example 143 of calculating the range of a matrix.

$$\text{ran}\begin{pmatrix}1&2&0&1\\1&2&1&2\\0&0&1&1\end{pmatrix} := \left\{\begin{pmatrix}1&2&0&1\\1&2&1&2\\0&0&1&1\end{pmatrix}\begin{pmatrix}x\\y\\z\\w\end{pmatrix} \mid \begin{pmatrix}x\\y\\z\\w\end{pmatrix} \in \mathbb{R}^4\right\}$$

$$= \left\{x\begin{pmatrix}1\\1\\0\end{pmatrix} + y\begin{pmatrix}2\\2\\0\end{pmatrix} + z\begin{pmatrix}0\\1\\1\end{pmatrix} + w\begin{pmatrix}1\\2\\1\end{pmatrix}\middle| x,y,z,w \in \mathbb{R}\right\}.$$

That is

$$\text{ran}\begin{pmatrix}1&2&0&1\\1&2&1&2\\0&0&1&1\end{pmatrix} = \text{span}\left\{\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}2\\2\\0\end{pmatrix},\begin{pmatrix}0\\1\\1\end{pmatrix},\begin{pmatrix}1\\2\\1\end{pmatrix}\right\}$$

but since

$$\text{RREF}\begin{pmatrix}1&2&0&1\\1&2&1&2\\0&0&1&1\end{pmatrix} = \begin{pmatrix}1&2&0&1\\0&0&1&1\\0&0&0&0\end{pmatrix}$$

the second and fourth columns (which are the non-pivot columns), can be expressed as linear combinations of columns to their left. They can then be removed from the set in the span to obtain

$$\text{ran}\begin{pmatrix}1&2&0&1\\1&2&1&2\\0&0&1&1\end{pmatrix} = \text{span}\left\{\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}0\\1\\1\end{pmatrix}\right\}.$$

It might occur to you that the range of the 3×4 matrix from the last example can be expressed as the range of a 3×2 matrix;

$$\operatorname{ran}\begin{pmatrix}1&2&0&1\\1&2&1&2\\0&0&1&1\end{pmatrix} = \operatorname{ran}\begin{pmatrix}1&0\\1&1\\0&1\end{pmatrix}.$$

Indeed, because the span of a set of vectors does not change when we replace the vectors with another set through an invertible process, we can calculate ranges through strings of equalities of ranges of matrices that differer by Elementary Column Operations, ECOs, ending with the range of a matrix in Column Reduced Echelon Form, CREF, with its zero columns deleted.

Example 144 Calculating a range with ECOs

$$\operatorname{ran}\begin{pmatrix}0&1&1\\1&3&1\\1&2&0\end{pmatrix} \overset{c_1\leftrightarrow c_3}{=} \operatorname{ran}\begin{pmatrix}1&1&0\\1&3&1\\0&2&1\end{pmatrix} \overset{c_2'=c_2-c_1}{=} \operatorname{ran}\begin{pmatrix}1&0&0\\1&2&1\\0&2&1\end{pmatrix} \overset{c_2'=\frac{1}{2}c_2}{=} \operatorname{ran}\begin{pmatrix}1&0&0\\1&1&1\\0&1&1\end{pmatrix}$$

$$\overset{c_3'=c_3-c_2}{=} \operatorname{ran}\begin{pmatrix}1&0&0\\1&1&0\\0&1&0\end{pmatrix} = \operatorname{ran}\begin{pmatrix}1&0\\1&1\\0&1\end{pmatrix}.$$

This is an efficient way to compute and encode the range of a matrix.

16.2 Image

Definition For any subset U of the domain S of a function $f : S \to T$ the **image** of U is

$$f(U) = \operatorname{Im} U := \{f(x) | x \in U\}.$$

Example 145 The image of the cube

$$U = \left\{ a\begin{pmatrix}1\\0\\0\end{pmatrix} + b\begin{pmatrix}0\\1\\0\end{pmatrix} + c\begin{pmatrix}0\\0\\1\end{pmatrix} \middle| a, b, c \in [0,1] \right\}$$

under multiplication by the matrix

$$M = \begin{pmatrix}1&0&0\\1&1&1\\0&0&1\end{pmatrix}$$

is the parallelepiped

$$\operatorname{Img} U = \left\{ a \begin{pmatrix}1\\1\\0\end{pmatrix} + b \begin{pmatrix}0\\1\\0\end{pmatrix} + c \begin{pmatrix}0\\1\\1\end{pmatrix} \middle| a, b, c \in [0,1] \right\}.$$

Note that for most subsets U of the domain S of a function f the image of U is not a vector space. The range of a function is the particular case of the image where the subset of the domain is the entire domain; $\operatorname{ran} f = \operatorname{Img} S$. For this reason, the range of f is also sometimes called the image of f and is sometimes denoted $\operatorname{im}(f)$ or $f(S)$. We have seen that the range of a matrix is always a span of vectors, and hence a vector space.

Note that we prefer the phrase "range of f" to the phrase "image of f" because we wish to avoid confusion between homophones; the word "image" is also used to describe a single element of the codomain assigned to a single element of the domain. For example, one might say of the function $A : \mathbb{R} \to \mathbb{R}$ with rule of correspondence $A(x =) = 2x - 1$ for all x in $\mathbb{R}$ that the image of 2 is 3 with this second meaning of the word "image" in mind. By contrast, one would never say that the range of 2 is 3 since the former is not a function and the latter is not a set.

For thinking about inverses of function we want to think in the oposite direction in a sense.

Definition The **pre-image** of any subset $U \subset T$ is

$$f^{-1}(U) := \{s \in S | f(s) \in U\} \subset S.$$

The pre-image of a set U is the set of all elements of S which map to U.

Example 146 The pre-image of the set $U = \left\{ a \begin{pmatrix}2\\1\\1\end{pmatrix} \middle| a \in [0,1] \right\}$ (a line segment) under the matrix

$$M = \begin{pmatrix}1 & 0 & 1\\0 & 1 & 1\\0 & 1 & 1\end{pmatrix} : \mathbb{R}^3 \to \mathbb{R}^3$$

is the set

$$\begin{aligned} M^{-1}U &= \{x | Mx = v \text{ for some } v \in U\} \\ &= \left\{ \begin{pmatrix}x\\y\\z\end{pmatrix} \middle| \begin{pmatrix}1 & 0 & 1\\0 & 1 & 1\\0 & 1 & 1\end{pmatrix} \begin{pmatrix}x\\y\\z\end{pmatrix} = a \begin{pmatrix}2\\1\\1\end{pmatrix} \text{ for some } a \in [0,1] \right\}. \end{aligned}$$

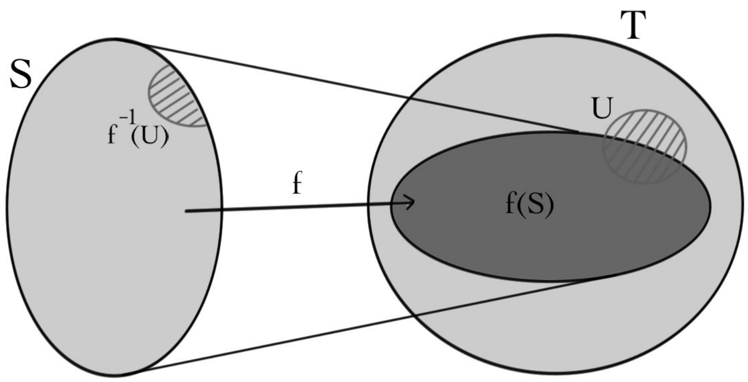


Figure 16.1: For the function $f : S \to T$, S is the domain, T is the target/codomain, $f(S)$ is the range and $f^{-1}(U)$ is the pre-image of $U \subset T$.

Since

$$\text{RREF}\left(\begin{array}{ccc|c}1 & 0 & 1 & 2a\\ 0 & 1 & 1 & a\\ 0 & 1 & 1 & a\end{array}\right) = \left(\begin{array}{ccc|c}1 & 0 & 1 & 2a\\ 0 & 1 & 1 & a\\ 0 & 0 & 0 & 0\end{array}\right)$$

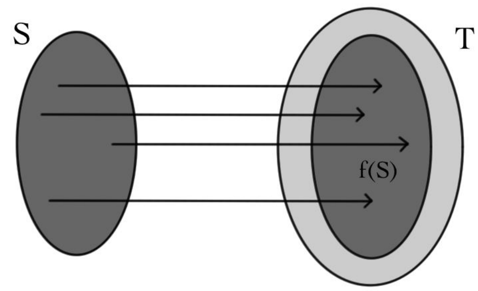
we have

$$M^{-1}U = \left\{ a\begin{pmatrix}2\\1\\0\end{pmatrix} + b\begin{pmatrix}-1\\-1\\1\end{pmatrix} \middle| a \in [0,1], b \in \mathbb{R} \right\},$$

a strip from a plane in $\mathbb{R}^3$.

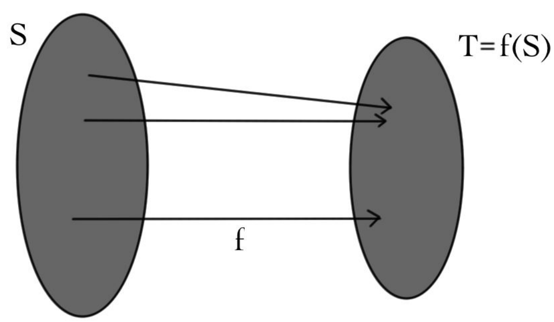
16.2.1 One-to-one and Onto

The function f is **one-to-one** (sometimes denoted 1:1) if different elements in S always map to different elements in T. That is, f is one-to-one if for any elements $x \neq y \in S$, we have that $f(x) \neq f(y)$, as pictured below.

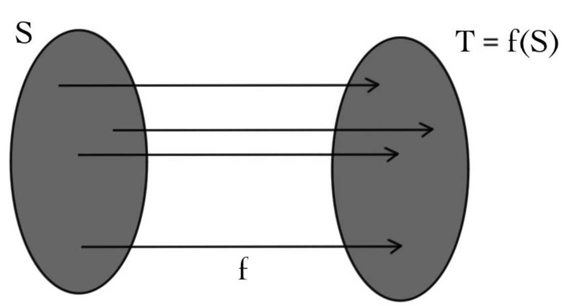


One-to-one functions are also called **injective** functions (and sometimes called monomorphisms.) Notice that injectivity is a condition on the pre-images of f.

The function f is **onto** if every element of T is mapped to by some element of S. That is, f is onto if for any $t \in T$, there exists some $s \in S$ such that $f(s) = t$. Onto functions are also called **surjective** functions (and sometimes epimorphisms.) Notice that surjectivity is a condition on the range of f.



If f is both injective and surjective, it is **bijective** (or an isomorphism.)



Theorem 16.2.1. *A function $f\colon S \to T$ has an inverse function $g\colon T \to S$ if and only if f is bijective.*

Proof. This is an "if and only if" statement so the proof has two parts.

1. *(Existence of an inverse $\Rightarrow$ bijective.)*

   Suppose that f has an inverse function g. We need to show f is bijective, which we break down into injective and surjective.

   - The function f is injective: Suppose that we have $s, s' \in S$ such that $f(s) = f(s')$. We must have that $g(f(s)) = s$ for any $s \in S$, so in particular $g(f(s)) = s$ and $g(f(s')) = s'$. But since $f(s) = f(s')$, we have $g(f(s)) = g(f(s'))$ so $s = s'$. Therefore, f is injective.
   - The function f is surjective: Let t be any element of T. We must have that $f(g(t)) = t$. Thus, $g(t)$ is an element of S which maps to t. So f is surjective.

2. *(Bijectivity $\Rightarrow$ existence of an inverse.)* Suppose that f is bijective. Hence f is surjective, so every element $t \in T$ has at least one pre-image. Being bijective, f is also injective, so every t has no more than one pre-image. Therefore, to construct an inverse function g, we simply define $g(t)$ to be the unique pre-image $f^{-1}(t)$ of t.

□

Now let us specialize to functions f that are linear maps between two vector spaces. Everything we said above for arbitrary functions is exactly the same for linear functions. However, the structure of vector spaces lets us say much more about one-to-one and onto functions whose domains are vector spaces than we can say about functions on general sets. For example, we know that a linear function always sends 0_V to 0_W, *i.e.*,

$$\boxed{f(0_V) = 0_W}$$

In Review Exercise 2, you will show that a linear transformation is one-to-one if and only if 0_V is the only vector that is sent to 0_W. Linear functions are unlike arbitrary functions between sets in that, by looking at *just one* (very special) vector, we can figure out whether f is one-to-one!

16.2.2 Kernel

Let $L\colon V \to W$ be a linear transformation. Suppose L is *not* injective. Then we can find $v_1 \neq v_2$ such that $Lv_1 = Lv_2$. So $v_1 - v_2 \neq 0$, but

$$L(v_1 - v_2) = 0.$$

Definition If $L\colon V \to W$ is a linear function then the set

$$\ker L = \{v \in V \mid Lv = 0_W\} \subset V$$

is called the **kernel of** L.

Notice that if L has matrix M in some basis, then finding the kernel of L is equivalent to solving the homogeneous system

$$MX = 0.$$

Example 147 Let $L(x, y) = (x + y, x + 2y, y)$. Is L one-to-one?

To find out, we can solve the linear system:

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right).$$

Then all solutions of $MX = 0$ are of the form $x = y = 0$. In other words, $\ker L = \{0\}$, and so L is injective.

Reading homework: problem 1

Notice that in the above example we found

$$\ker \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} = \ker \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

In general, an efficient way to get the kernel of a matrix is to write a string of equalities between kernels of matrices which differ by row operations and, once RREF is reached, note that the linear relationships between the columns for a basis for the nullspace.

Example 148 of calculating the kernel of a matrix.

$$\ker\begin{pmatrix}1&2&0&1\\1&2&1&2\\0&0&1&1\end{pmatrix}=\ker\begin{pmatrix}1&2&0&1\\0&0&1&1\\0&0&1&1\end{pmatrix}=\ker\begin{pmatrix}1&2&0&1\\0&0&1&1\\0&0&0&0\end{pmatrix}$$

$$=\operatorname{span}\left\{\begin{pmatrix}-2\\1\\0\\0\end{pmatrix},\begin{pmatrix}-1\\0\\-1\\1\end{pmatrix}\right\}.$$

The two column vectors in this last line describe linear relations between the columns c_1, c_2, c_3, c_4. In particular $-2c_1 + 1c_2 = 0$ and $-c_1 - c_3 + c_4 = 0$.

In general, a description of the kernel of a matrix should be of the form $\operatorname{span}\{v_1, v_2, \ldots, v_n\}$ with one vector v_i for each non-pivot column. To agree with the standard procedure, think about how to describe each non-pivot column in terms of columns to its left; this will yield an expression of the form wherein each vector has a 1 as its last non-zero entry. (Think of Column Reduced Echelon Form, CREF.)

Thinking again of augmented matrices, if a matrix has more than one element in its kernel then it is not invertible since the existence of multiple solutions to $Mx = 0$ implies that $\operatorname{RREF} M \neq I$. However just because the kernel of a linear function is trivial does not mean that the function is invertible.

Example 149 $\ker\begin{pmatrix}1&0\\1&1\\0&1\end{pmatrix}=\left\{\begin{pmatrix}0\\0\end{pmatrix}\right\}$ since the matrix has no non-pivot columns.

However, $\begin{pmatrix}1&0\\1&1\\0&1\end{pmatrix}:\mathbb{R}^2\to\mathbb{R}^3$ is not invertible because there are many things in its codomain that are not in its range, such as $\begin{pmatrix}1\\0\\0\end{pmatrix}$.

A trivial kernel only gives us half of what is needed for invertibility.

Theorem 16.2.2. *A linear transformation $L\colon V\to W$ is injective iff*

$$\ker L = \{0_V\}\,.$$

Proof. The proof of this theorem is Review Exercise 2. □

Theorem 16.2.3. *If* $L\colon V \xrightarrow{\text{linear}} W$ *then* $\ker L$ *is a subspace of* V*.*

Proof. Notice that if $L(v) = 0$ and $L(u) = 0$, then for any constants c, d, $L(cu+dv) = 0$. Then by the subspace theorem, the kernel of L is a subspace of V. □

Example 150 Let $L\colon \mathbb{R}^3 \to \mathbb{R}$ be the linear transformation defined by $L(x, y, z) = (x + y + z)$. Then $\ker L$ consists of all vectors $(x, y, z) \in \mathbb{R}^3$ such that $x + y + z = 0$. Therefore, the set

$$V = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}$$

is a subspace of $\mathbb{R}^3$.

When $L : V \to V$, the above theorem has an interpretation in terms of the eigenspaces of L. Suppose L has a zero eigenvalue. Then the associated eigenspace consists of all vectors v such that $Lv = 0v = 0$; the 0-eigenspace of L is exactly the kernel of L.

In the example where $L(x, y) = (x + y, x + 2y, y)$, the map L is clearly not surjective, since L maps $\mathbb{R}^2$ to a plane through the origin in $\mathbb{R}^3$. But any plane through the origin is a subspace. In general notice that if $w = L(v)$ and $w' = L(v')$, then for any constants c, d, linearity of L ensures that

$$cw + dw' = L(cv + dv')\,.$$

Now the subspace theorem strikes again, and we have the following theorem:

Theorem 16.2.4. *If* $L\colon V \to W$ *is linear then the range* $L(V)$ *is a subspace of* W*.*

Example 151 Let $L(x, y) = (x + y, x + 2y, y)$. The range of L is a plane through the origin and thus a subspace of $\mathbb{R}^3$. Indeed the matrix of L in the standard basis is

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix}.$$

The columns of this matrix encode the possible outputs of the function L because

$$L(x, y) = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

Thus

$$L(\mathbb{R}^2) = \operatorname{span}\left\{ \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\2\\1\end{pmatrix} \right\}$$

Hence, when bases and a linear transformation is are given, people often refer to its range as the *column space* of the corresponding matrix.

To find a basis of the range of L, we can start with a basis $S = \{v_1, \ldots, v_n\}$ for V. Then the most general input for L is of the form $\alpha^1 v_1 + \cdots + \alpha^n v_n$. In turn, its most general output looks like

$$L\big(\alpha^1 v_1 + \cdots + \alpha^n v_n\big) = \alpha^1 L v_1 + \cdots + \alpha^n L v_n \in \operatorname{span}\{Lv_1, \ldots\ Lv_n\}\,.$$

Thus

$$L(V) = \operatorname{span} L(S) = \operatorname{span}\{Lv_1, \ldots, Lv_n\}\,.$$

However, the set $\{Lv_1, \ldots, Lv_n\}$ may not be linearly independent; we must solve

$$c^1 L v_1 + \cdots + c^n L v_n = 0\,,$$

to determine whether it is. By finding relations amongst the elements of $L(S) = \{Lv_1, \ldots, Lv_n\}$, we can discard vectors until a basis is arrived at. The size of this basis is the dimension of the range of L, which is known as the *rank* of L.

Definition The **rank** of a linear transformation L is the dimension of its range. The **nullity** of a linear transformation is the dimension of the kernel.

The notation for these numbers is

$$\boxed{\operatorname{null} L := \dim \ker L,}$$

$$\boxed{\operatorname{rank} L := \dim L(V) = \dim\ \operatorname{ran} L.}$$

Theorem 16.2.5 (Dimension Formula). *Let $L: V \to W$ be a linear transformation, with V a finite-dimensional vector space*[1]*. Then:*

$$\begin{aligned} \dim V &= \dim \ker V + \dim L(V) \\ &= \operatorname{null} L + \operatorname{rank} L. \end{aligned}$$

Proof. Pick a basis for V:

$$\{v_1, \ldots, v_p, u_1, \ldots, u_q\},$$

where $v_1, \ldots, v_p$ is also a basis for $\ker L$. This can always be done, for example, by finding a basis for the kernel of L and then extending to a basis for V. Then $p = \operatorname{null} L$ and $p + q = \dim V$. Then we need to show that $q = \operatorname{rank} L$. To accomplish this, we show that $\{L(u_1), \ldots, L(u_q)\}$ is a basis for $L(V)$.

To see that $\{L(u_1), \ldots, L(u_q)\}$ spans $L(V)$, consider any vector w in $L(V)$. Then we can find constants c^i, d^j such that:

$$\begin{aligned} w &= L(c^1 v_1 + \cdots + c^p v_p + d^1 u_1 + \cdots + d^q u_q) \\ &= c^1 L(v_1) + \cdots + c^p L(v_p) + d^1 L(u_1) + \cdots + d^q L(u_q) \\ &= d^1 L(u_1) + \cdots + d^q L(u_q) \text{ since } L(v_i) = 0, \\ \Rightarrow L(V) &= \operatorname{span}\{L(u_1), \ldots, L(u_q)\}. \end{aligned}$$

Now we show that $\{L(u_1), \ldots, L(u_q)\}$ is linearly independent. We argue by contradiction. Suppose there exist constants d^j (not all zero) such that

$$\begin{aligned} 0 &= d^1 L(u_1) + \cdots + d^q L(u_q) \\ &= L(d^1 u_1 + \cdots + d^q u_q). \end{aligned}$$

But since the u^j are linearly independent, then $d^1 u_1 + \cdots + d^q u_q \neq 0$, and so $d^1 u_1 + \cdots + d^q u_q$ is in the kernel of L. But then $d^1 u_1 + \cdots + d^q u_q$ must be in the span of $\{v_1, \ldots, v_p\}$, since this was a basis for the kernel. This contradicts the assumption that $\{v_1, \ldots, v_p, u_1, \ldots, u_q\}$ was a basis for V, so we are done. □

[1] The formula still makes sense for infinite dimensional vector spaces, such as the space of all polynomials, but the notion of a basis for an infinite dimensional space is more sticky than in the finite-dimensional case. Furthermore, the dimension formula for infinite dimensional vector spaces isn't useful for computing the rank of a linear transformation, since an equation like $\infty = \infty + x$ cannot be solved for x. As such, the proof presented assumes a finite basis for V.

Reading homework: problem 2

Example 152 (Row rank equals column rank)
Suppose M is an $m \times n$ matrix. The matrix M itself is a linear transformation $M : \mathbb{R}^n \to \mathbb{R}^m$ but it must also be the matrix of some linear transformation

$$L : V \stackrel{\text{linear}}{\longrightarrow} W\,.$$

Here we only know that $\dim V = n$ and $\dim W = m$. The rank of the map L is the dimension of its image and also the number of linearly independent columns of M. Hence, this is sometimes called the *column rank* of M. The dimension formula predicts the dimension of the kernel, *i.e.* the nullity: $\operatorname{null} L = \dim V - \operatorname{rank} L = n - r$.

To compute the kernel we would study the linear system

$$Mx = 0\,,$$

which gives m equations for the n-vector x. The *row rank* of a matrix is the number of linearly independent rows (viewed as vectors). Each linearly independent row of M gives an independent equation satisfied by the n-vector x. Every independent equation on x reduces the size of the kernel by one, so if the row rank is s, then $\operatorname{null} L + s = n$. Thus we have two equations:

$$\operatorname{null} L + s = n \text{ and } \operatorname{null} L = n - r\,.$$

From these we conclude the $r = s$. In other words, the row rank of M equals its column rank.

16.3 Summary

We have seen that a linear transformation has an inverse if and only if it is bijective (*i.e.*, one-to-one and onto). We also know that linear transformations can be represented by matrices, and we have seen many ways to tell whether a matrix is invertible. Here is a list of them:

Theorem 16.3.1 (Invertibility)**.** *Let V be an n-dimensional vector space and suppose $L : V \to V$ is a linear transformation with matrix M in some basis. Then M is an $n \times n$ matrix, and the following statements are equivalent:*

1. *If v is any vector in $\mathbb{R}^n$, then the system $Mx = v$ has exactly one solution.*
2. *The matrix M is row-equivalent to the identity matrix.*
3. *If v is any vector in V, then $L(x) = v$ has exactly one solution.*
4. *The matrix M is invertible.*
5. *The homogeneous system $Mx = 0$ has no non-zero solutions.*
6. *The determinant of M is not equal to 0.*
7. *The transpose matrix M^T is invertible.*
8. *The matrix M does not have 0 as an eigenvalue.*
9. *The linear transformation L does not have 0 as an eigenvalue.*
10. *The characteristic polynomial $\det(\lambda I - M)$ does not have 0 as a root.*
11. *The columns (or rows) of M span $\mathbb{R}^n$.*
12. *The columns (or rows) of M are linearly independent.*
13. *The columns (or rows) of M are a basis for $\mathbb{R}^n$.*
14. *The linear transformation L is injective.*
15. *The linear transformation L is surjective.*
16. *The linear transformation L is bijective.*

Note: it is important that M be an $n \times n$ matrix! If M is not square, then it can't be invertible, and many of the statements above are no longer equivalent to each other.

Proof. Many of these equivalences were proved earlier in other chapters. Some were left as review questions or sample final questions. The rest are left as exercises for the reader. □



`Invertibility Conditions`



16.4 Review Problems

Webwork:

Reading Problems	1, 2,
Elements of kernel	3
Basis for column space	4
Basis for kernel	5
Basis for kernel and range	6
Orthonomal range basis	7
Orthonomal kernel basis	8
Orthonomal kernel and range bases	9
Orthonomal kernel, range and row space bases	10
Rank	11

1. Consider an arbitrary matrix $M : \mathbb{R}^m \to \mathbb{R}^n$.

   (a) Argue that $Mx = 0$ if only if x is perpendicular to all columns of M^T.

   (b) Argue that $Mx = 0$ if only if x is perpendicular to all of the linear combinations of the columns of M^T.

   (c) Argue that $\ker M$ is perpendicular to $\operatorname{ran} M^T$.

   (d) Argue further $\mathbb{R}^m = \ker M \oplus \operatorname{ran} M^T$.

   (e) Argue analogously that $\mathbb{R}^n = \ker M^T \oplus \operatorname{ran} M$.

   The equations in the last two parts describe how a linear transformation $M : \mathbb{R}^m \to \mathbb{R}^n$ determines orthogonal decompositions of both it's domain and target. This result sometimes goes by the humble name **The Fundamental Theorem of Linear Algebra**.

2. Let $L\colon V \to W$ be a linear transformation. Show that $\ker L = \{0_V\}$ if and only if L is one-to-one:

   (a) *(Trivial kernel ⇒ injective.)* Suppose that $\ker L = \{0_V\}$. Show that L is one-to-one. Think about methods of proof–does a proof by contradiction, a proof by induction, or a direct proof seem most appropriate?

   (b) *(Injective ⇒ trivial kernel.)* Now suppose that L is one-to-one. Show that $\ker L = \{0_V\}$. That is, show that 0_V is in $\ker L$, and then show that there are no other vectors in $\ker L$.

Hint

3. Let $\{v_1, \ldots, v_n\}$ be a basis for V and $L : V \to W$ is a linear function. Carefully explain why

$$L(V) = \operatorname{span}\{Lv_1, \ldots, Lv_n\}\,.$$

4. Suppose $L \colon \mathbb{R}^4 \to \mathbb{R}^3$ whose matrix M in the standard basis is row equivalent to the following matrix:

$$\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} = \operatorname{RREF}(M) \sim M.$$

   (a) *Explain* why the first three columns of the original matrix M form a basis for $L(\mathbb{R}^4)$.

   (b) *Find and describe* an algorithm (*i.e.*, a general procedure) for computing a basis for $L(\mathbb{R}^n)$ when $L \colon \mathbb{R}^n \to \mathbb{R}^m$.

   (c) *Use* your algorithm to find a basis for $L(\mathbb{R}^4)$ when $L \colon \mathbb{R}^4 \to \mathbb{R}^3$ is the linear transformation whose matrix M in the standard basis is

$$\begin{pmatrix} 2 & 1 & 1 & 4 \\ 0 & 1 & 0 & 5 \\ 4 & 1 & 1 & 6 \end{pmatrix}.$$

5. Claim:

   If $\{v_1, \ldots, v_n\}$ is a basis for $\ker L$, where $L \colon V \to W$, then it is always possible to extend this set to a basis for V.

   Choose some simple yet non-trivial linear transformations with non-trivial kernels and verify the above claim for those transformations.

6. Let $P_n(x)$ be the space of polynomials in x of degree less than or equal to n, and consider the derivative operator

$$\frac{d}{dx} : P_n(x) \to P_n(x)\,.$$

Find the dimension of the kernel and image of this operator. What happens if the target space is changed to $P_{n-1}(x)$ or $P_{n+1}(x)$?

Now consider $P_2(x, y)$, the space of polynomials of degree two or less in x and y. (Recall how degree is counted; xy is degree two, y is degree one and x^2y is degree three, for example.) Let

$$L := \frac{\partial}{\partial x} + \frac{\partial}{\partial y} : P_2(x, y) \to P_2(x, y).$$

(For example, $L(xy) = \frac{\partial}{\partial x}(xy) + \frac{\partial}{\partial y}(xy) = y + x$.) Find a basis for the kernel of L. Verify the dimension formula in this case.

7. Lets demonstrate some ways the dimension formula can break down if a vector space is infinite dimensional.

   (a) Let $\mathbb{R}[x]$ be the vector space of all polynomials in the variable x with real coefficients. Let $D = \frac{d}{dx}$ be the usual derivative operator. Show that the range of D is $\mathbb{R}[x]$. What is $\ker D$?

   Hint: Use the basis $\{x^n \mid n \in \mathbb{N}\}$.

   (b) Let $L\colon \mathbb{R}[x] \to \mathbb{R}[x]$ be the linear map

   $$L(p(x)) = xp(x)\,.$$

   What is the kernel and range of M?

   (c) Let V be an infinite dimensional vector space and $L\colon V \to V$ be a linear operator. Suppose that $\dim\ker L < \infty$, show that $\dim L(V)$ is infinite. Also show that when $\dim L(V) < \infty$ that $\dim\ker L$ is infinite.

8. This question will answer the question, "If I choose a bit vector *at random*, what is the probability that it lies in the span of some other vectors?"

   i. Given a collection S of k bit vectors in B^3, consider the bit matrix M whose columns are the vectors in S. Show that S is linearly independent if and only if the kernel of M is trivial, namely the set $\ker M = \{v \in B^3 \mid Mv = 0\}$ contains only the zero vector.

ii. Give some method for choosing a random bit vector v in B^3. Suppose S is a collection of 2 linearly independent bit vectors in B^3. How can we tell whether $S \cup \{v\}$ is linearly independent? Do you think it is likely or unlikely that $S \cup \{v\}$ is linearly independent? Explain your reasoning.

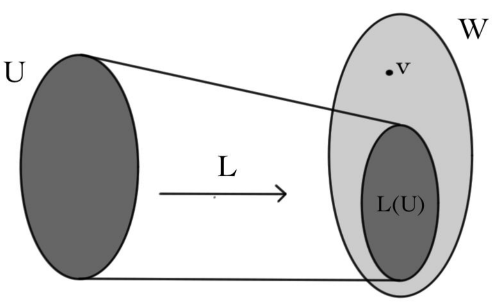
iii. If P is the characteristic polynomial of a 3×3 bit matrix, what must the degree of P be? Given that each coefficient must be either 0 or 1, how many possibilities are there for P? How many of these possible characteristic polynomials have 0 as a root? If M is a 3×3 bit matrix chosen at random, what is the probability that it has 0 as an eigenvalue? (Assume that you are choosing a random matrix M in such a way as to make each characteristic polynomial equally likely.) What is the probability that the columns of M form a basis for B^3? (Hint: what is the relationship between the kernel of M and its eigenvalues?)

Note: We could ask the same question for real vectors: If I choose a real vector at random, what is the probability that it lies in the span of some other vectors? In fact, once we write down a reasonable way of choosing a random real vector, if I choose a real vector in $\mathbb{R}^n$ at random, the probability that it lies in the span of $n-1$ other real vectors is zero!

17

Least squares and Singular Values

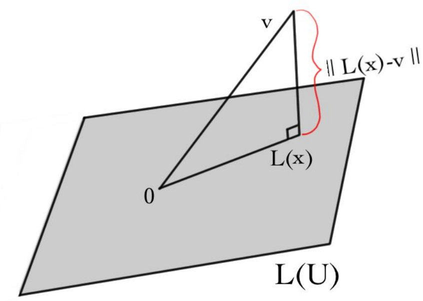
Consider the linear algebraic equation $L(x) = v$, where $L\colon U \xrightarrow{\text{linear}} W$ and $v \in W$ are known while x is unknown. As we have seen, this system may have one solution, no solutions, or infinitely many solutions. But if v is not in the range of L there will *never* be any solutions for $L(x) = v$.



However, for many applications we do not need an exact solution of the system; instead, we may only need the best approximation possible.

> "My work always tried to unite the Truth with the Beautiful, but when I had to choose one or the other, I usually chose the Beautiful."
>
> – Hermann Weyl.

If the vector space W has a notion of lengths of vectors, we can try to find x that minimizes $||L(x) - v||$.



This method has many applications, such as when trying to fit a (perhaps linear) function to a "noisy" set of observations. For example, suppose we measured the position of a bicycle on a racetrack once every five seconds. Our observations won't be exact, but so long as the observations are right on average, we can figure out a best-possible linear function of position of the bicycle in terms of time.

Suppose M is the matrix for the linear function $L : U \to W$ in some bases for U and W. The vectors v and x are represented by column vectors V and X in these bases. Then we need to approximate

$$MX - V \approx 0\,.$$

Note that if $\dim U = n$ and $\dim W = m$ then M can be represented by an $m \times n$ matrix and x and v as vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Thus, we can write $W = L(U) \oplus L(U)^{\perp}$. Then we can uniquely write $v = v^{\parallel} + v^{\perp}$, with $v^{\parallel} \in L(U)$ and $v^{\perp} \in L(U)^{\perp}$.

Thus we should solve $L(u) = v^{\parallel}$. In components, $v^{\perp}$ is just $V - MX$, and is the part we will eventually wish to minimize.

In terms of M, recall that $L(V)$ is spanned by the columns of M. (In the standard basis, the columns of M are Me_1, …, Me_n.) Then $v^{\perp}$ must be perpendicular to the columns of M. *i.e.*, $M^T(V - MX) = 0$, or

$$M^T MX = M^T V.$$

Solutions of $M^T MX = M^T V$ for X are called *least squares* solutions to $MX = V$. Notice that any solution X to $MX = V$ is a least squares solution.

However, the converse is often false. In fact, the equation $MX = V$ may have no solutions at all, but still have least squares solutions to $M^T MX = M^T V$.

Observe that since M is an $m \times n$ matrix, then M^T is an $n \times m$ matrix. Then $M^T M$ is an $n \times n$ matrix, and is symmetric, since $(M^T M)^T = M^T M$. Then, for any vector X, we can evaluate $X^T M^T MX$ to obtain a number. This is a very nice number, though! It is just the length $|MX|^2 = (MX)^T(MX) = X^T M^T MX$.

Reading homework: problem 1

Now suppose that $\ker L = \{0\}$, so that the only solution to $MX = 0$ is $X = 0$. (This need not mean that M is invertible because M is an $n \times m$ matrix, so not necessarily square.) However the square matrix $M^T M$ *is* invertible. To see this, suppose there was a vector X such that $M^T MX = 0$. Then it would follow that $X^T M^T MX = |MX|^2 = 0$. In other words the vector MX would have zero length, so could only be the zero vector. But we are assuming that $\ker L = \{0\}$ so $MX = 0$ implies $X = 0$. Thus the kernel of $M^T M$ is $\{0\}$ so this matrix is invertible. So, in this case, the least squares solution (the X that solves $M^T MX = MV$) is unique, and is equal to

$$X = (M^T M)^{-1} M^T V.$$

In a nutshell, this is the least squares method:

- Compute $M^T M$ and $M^T V$.
- Solve $(M^T M)X = M^T V$ by Gaussian elimination.

Example 153 Captain Conundrum falls off of the leaning tower of Pisa and makes three (rather shaky) measurements of his velocity at three different times.

t s	v m/s
1	11
2	19
3	31

Having taken some calculus[1], he believes that his data are best approximated by a straight line

$$v = at + b.$$

[1]In fact, he is a *Calculus Superhero*.

Then he should find a and b to best fit the data.

$$\begin{aligned} 11 &= a \cdot 1 + b \\ 19 &= a \cdot 2 + b \\ 31 &= a \cdot 3 + b. \end{aligned}$$

As a system of linear equations, this becomes:

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} 11 \\ 19 \\ 31 \end{pmatrix}.$$

There is likely no actual straight line solution, so instead solve $M^T M X = M^T V$.

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 11 \\ 19 \\ 31 \end{pmatrix}.$$

This simplifies to

$$\left(\begin{array}{cc|c} 14 & 6 & 142 \\ 6 & 3 & 61 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 0 & 10 \\ 0 & 1 & \frac{1}{3} \end{array}\right).$$

Thus, the least-squares fit is the line

$$v = 10\, t + \frac{1}{3}.$$

Notice that this equation implies that Captain Conundrum accelerates towards Italian soil at 10 m/s^2 (which is an excellent approximation to reality) and that he started at a downward velocity of $\frac{1}{3}$ m/s (perhaps somebody gave him a shove...)!

17.1 Projection Matrices

We have seen that even if $MX = V$ has no solutions $M^T M X = M^T V$ does have solutions. One way to think about this is, since the codomain of M is the direct sum

$$\text{codom M} = \text{ran} M \oplus \ker M^T$$

there is a unique way to write $V = V_r + V_k$ with $V_k \in \ker M^T$ and $V_r \in \text{ran }\ M$, and it is clear that $Mx = V$ only has a solution of $V \in \text{ran}\, M \Leftrightarrow V_k = 0$. If not, then the closest thing to a solution of $MX = V$ is a solution to

$MX = V_r$. We learned to find solutions to this in the previous subsection of this book.

But here is another question, how can we determine what V_r is given M and V? The answer is simple; suppose X is a solution to $MX = V_r$. Then

$$MX = V_r \implies M^T Mx = M^T V_r \implies M^T Mx = M^T(V_r + 0)$$

$$\implies M^T Mx = M^T(V_r + V_k) \implies M^T Mx = M^T V \implies X = (M^T M)^{-1} M^T V$$

if indeed $M^T M$ is invertible. Since, by assumption, X is a solution

$$\boxed{M(M^T M)^{-1} M^T V = V_r.}$$

That is, the matrix which projects V onto its $\operatorname{ran} M$ part is $M(M^T M)^{-1} M^T$.

Example 154 To project $\begin{pmatrix}1\\1\\1\end{pmatrix}$ onto span $\left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix}\right\}$ = ran $\begin{pmatrix}1 & 1\\1 & -1\\0 & 0\end{pmatrix}$ multiply by the matrix

$$\begin{pmatrix}1 & 1\\1 & -1\\0 & 0\end{pmatrix}\left[\begin{pmatrix}1 & 1 & 0\\1 & -1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\1 & -1\\0 & 0\end{pmatrix}\right]^{-1}\begin{pmatrix}1 & 1 & 0\\1 & -1 & 0\end{pmatrix}$$

$$= \begin{pmatrix}1 & 1\\1 & -1\\0 & 0\end{pmatrix}\begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}^{-1}\begin{pmatrix}1 & 1 & 0\\1 & -1 & 0\end{pmatrix}$$

$$= \frac{1}{2}\begin{pmatrix}1 & 1\\1 & -1\\0 & 0\end{pmatrix}\begin{pmatrix}1 & 1 & 0\\1 & -1 & 0\end{pmatrix} = \frac{1}{2}\begin{pmatrix}2 & 0 & 0\\0 & 2 & 0\\0 & 0 & 0\end{pmatrix}.$$

This gives

$$\frac{1}{2}\begin{pmatrix}2 & 0 & 0\\0 & 2 & 0\\0 & 0 & 0\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\1\\0\end{pmatrix}.$$

17.2 Singular Value Decomposition

Suppose

$$L : V \xrightarrow{\text{linear}} W\, .$$

It is unlikely that $\dim V =: n = m := \dim W$ so a $m \times n$ matrix M of L in bases for V and W will not be square. Therefore there is no eigenvalue problem we can use to uncover a preferred basis. However, if the vector spaces V and W both have inner products, there does exist an analog of the eigenvalue problem, namely the singular values of L.

Before giving the details of the powerful technique known as the singular value decomposition, we note that it is an excellent example of what Eugene Wigner called the "Unreasonable Effectiveness of Mathematics":

> There is a story about two friends who were classmates in high school, talking about their jobs. One of them became a statistician and was working on population trends. He showed a reprint to his former classmate. The reprint started, as usual with the Gaussian distribution and the statistician explained to his former classmate the meaning of the symbols for the actual population and so on. His classmate was a bit incredulous and was not quite sure whether the statistician was pulling his leg. "How can you know that?" was his query. "And what is this symbol here?" "Oh," said the statistician, this is "π." "And what is that?" "The ratio of the circumference of the circle to its diameter." "Well, now you are pushing your joke too far," said the classmate, "surely the population has nothing to do with the circumference of the circle."
>
> Eugene Wigner, Commun. Pure and Appl. Math. **XIII**, 1 (1960).

Whenever we mathematically model a system, any "canonical quantities" (those that do not depend on any choices we make for calculating them) will correspond to important features of the system. For examples, the eigenvalues of the eigenvector equation you found in review question 1, chapter 12 encode the notes and harmonics that a guitar string can play!

Singular values appear in many linear algebra applications, especially those involving very large data sets such as statistics and signal processing.

Let us focus on the $m \times n$ matrix M of a linear transformation $L : V \to W$ written in orthonormal bases for the input and outputs of L (notice, the existence of these othonormal bases is predicated on having inner products for V and W). Even though the matrix M is not square, both the matrices MM^T and M^TM are square and symmetric! In terms of linear transformations M^T is the matrix of a linear transformation

$$L^* : W \xrightarrow{\text{linear}} V\, .$$

Thus $LL^* : W \to W$ and $L^*L : V \to V$ and both have eigenvalue problems. Moreover, as is shown in Chapter 15, both L^*L and LL^* have orthonormal bases of eigenvectors, and both MM^T and M^TM can be diagonalized.

Next, let us make a simplifying assumption, namely $\ker L = \{0\}$. This is not necessary, but will make some of our computations simpler. Now suppose we have found an orthonormal basis $(u_1, \ldots, u_n)$ for V composed of eigenvectors for L^*L. That is

$$L^*Lu_i = \lambda_i u_i \, .$$

Then multiplying by L gives

$$LL^*Lu_i = \lambda_i Lu_i \, .$$

I.e., Lu_i is an eigenvector of LL^*. The vectors $(Lu_1, \ldots, Lu_n)$ are linearly independent, because $\ker L = \{0\}$ (this is where we use our simplifying assumption, but you can try and extend our analysis to the case where it no longer holds).

Lets compute the angles between and lengths of these vectors. For that we express the vectors u_i in the bases used to compute the matrix M of L. Denoting these column vectors by U_i we then compute

$$(MU_i) \cdot (MU_j) = U_i^T M^T M U_j = \lambda_j \, U_i^T U_j = \lambda_j \, U_i \cdot U_j = \lambda_j \delta_{ij} \, .$$

We see that vectors $(Lu_1, \ldots, Lu_n)$ are orthogonal but not orthonormal. Moreover, the length of Lu_i is $\sqrt{\lambda_i}$. Normalizing gives the orthonormal and linearly independent ordered set

$$\left(\frac{Lu_1}{\sqrt{\lambda_1}}, \ldots, \frac{Lu_n}{\sqrt{\lambda_n}} \right) .$$

In general, this cannot be a basis for W since $\ker L = \{0\}$, $\dim L(V) = \dim V$, and in turn $\dim V \leq \dim W$, so $n \leq m$.

However, it is a subset of the eigenvectors of LL^* so there is an orthonormal basis of eigenvectors of LL^* of the form

$$O' = \left(\frac{Lu_1}{\sqrt{\lambda_1}}, \ldots, \frac{Lu_n}{\sqrt{\lambda_n}}, v_{n+1}, \ldots, v_m \right) =: (v_1, \ldots, v_m) \, .$$

Now lets compute the matrix of L with respect to the orthonormal basis $O = (u_1, \ldots, u_n)$ for V and the orthonormal basis $O' = (v_1, \ldots, v_m)$ for W.

As usual, our starting point is the computation of L acting on the input basis vectors;

$$LO = \big(Lu_1, \ldots, Lu_n\big) = \big(\sqrt{\lambda_1}\, v_1, \ldots, \sqrt{\lambda_n}\, v_n\big)$$
$$= \big(v_1, \ldots, v_m\big) \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} .$$

The result is very close to diagonalization; the numbers $\sqrt{\lambda_i}$ along the leading diagonal are called the singular values of L.

Example 155 Let the matrix of a linear transformation be

$$M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -1 & 1 \\ -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} .$$

Clearly $\ker M = \{0\}$ while

$$M^T M = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{pmatrix}$$

which has eigenvalues and eigenvectors

$$\lambda = 1\,,\ u_1 := \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}; \qquad \lambda = 2\,,\ u_2 := \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} .$$

so our orthonormal input basis is

$$O = \left(\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right) .$$

These are called the *right singular vectors* of M. The vectors

$$Mu_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \text{ and } Mu_2 = \begin{pmatrix} 0 \\ -\sqrt{2} \\ 0 \end{pmatrix}$$

are eigenvectors of

$$MM^T = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 2 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

with eigenvalues 1 and 2, respectively. The third eigenvector (with eigenvalue 0) of MM^T is

$$v_3 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

The eigenvectors Mu_1 and Mu_2 are necessarily orthogonal, dividing them by their lengths we obtain the *left singular vectors* and in turn our orthonormal output basis

$$O' = \left(\begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right).$$

The new matrix M' of the linear transformation given by M with respect to the bases O and O' is

$$M' = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{pmatrix},$$

so the singular values are $1, \sqrt{2}$.

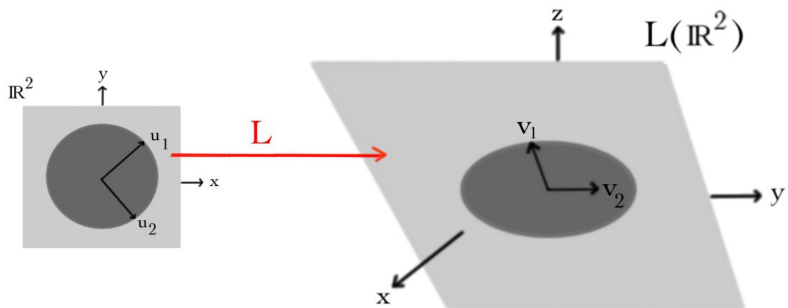
Finally note that arranging the column vectors of O and O' into change of basis matrices

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, \qquad Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & -1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix},$$

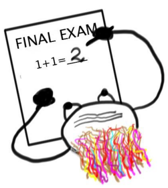
we have, as usual,

$$M' = Q^{-1}MP.$$

Singular vectors and values have a very nice geometric interpretation; they provide an orthonormal bases for the domain and range of L and give the factors by which L stretches the orthonormal input basis vectors. This is depicted below for the example we just computed.



Congratulations, you have reached the end of the book!



Now test your skills on the sample final exam.

17.3 Review Problems

Webwork: | Reading Problem | 1, |

1. Let $L : U \to V$ be a linear transformation. Suppose $v \in L(U)$ and you have found a vector $u_{\rm ps}$ that obeys $L(u_{\rm ps}) = v$.

   Explain why you need to compute $\ker L$ to describe the solution set of the linear system $L(u) = v$.

 Hint 

2. Suppose that M is an $m \times n$ matrix with trivial kernel. Show that for any vectors u and v in $\mathbb{R}^m$:

   - $u^T M^T M v = v^T M^T M u$.

- $v^T M^T M v \geq 0$. In case you are concerned (you don't need to be) and for future reference, the notation $v \geq 0$ means each component $v^i \geq 0$.
- If $v^T M^T M v = 0$, then $v = 0$.

(Hint: Think about the dot product in $\mathbb{R}^n$.)



Hint



3. Rewrite the Gram-Schmidt algorithm in terms of projection matrices.

4. Show that if $v_1, \ldots, v_k$ are linearly independent that the matrix $M = (v_1 \cdots v_k)$ is not necessarily invertible but the matrix $M^T M$ is invertible.

5. Write out the singular value decomposition theorem of a 3×1, a 3×2, and a 3×3 symmetric matrix. Make it so that none of the components of your matrices are zero but your computations are simple. Explain why you choose the matrices you choose.

6. Find the best polynomial approximation to a solution to the differential equation $\frac{d}{dx} f = x + x^2$ by considering the derivative to have domain and codomain span $\{1, x, x^2\}$.
(Hint: Begin by defining bases for the domain and codomain.)

A List of Symbols

$\in$	"Is an element of".
$\sim$	"Is equivalent to", see equivalence relations. Also, "is row equivalent to" for matrices.
$\mathbb{R}$	The real numbers.
I_n	The $n \times n$ identity matrix.
$P_n^{\mathbb{F}}$	The vector space of polynomials of degree at most n with coefficients in the field $\mathbb{F}$.
$\mathbb{M}_k^r$	The vector space of $r \times k$ matrices.

$\mathcal{B}$

Fields

Definition A *field* $\mathbb{F}$ is a set with two operations $+$ and $\cdot$, such that for all $a, b, c \in \mathbb{F}$ the following axioms are satisfied:

A1. Addition is associative $(a + b) + c = a + (b + c)$.

A2. There exists an additive identity 0.

A3. Addition is commutative $a + b = b + a$.

A4. There exists an additive inverse $-a$.

M1. Multiplication is associative $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

M2. There exists a multiplicative identity 1.

M3. Multiplication is commutative $a \cdot b = b \cdot a$.

M4. There exists a multiplicative inverse a^{-1} if $a \neq 0$.

D. The distributive law holds $a \cdot (b + c) = ab + ac$.

Roughly, all of the above mean that you have notions of $+$, $-$, $\times$ and $\div$ just as for regular real numbers.

Fields are a very beautiful structure; some examples are rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$, and complex numbers $\mathbb{C}$. These examples are infinite, however this does not necessarily have to be the case. The smallest

example of a field has just two elements, $\mathbb{Z}_2 = \{0, 1\}$ or *bits*. The rules for addition and multiplication are the usual ones save that

$$1 + 1 = 0 .$$

C

Online Resources

Here are some internet places to get linear algebra help:

- Strang's MIT Linear Algebra Course. Videos of lectures and more:

  http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/

- Beezer's online Linear Algebra Course

  http://linear.ups.edu/version3.html

- The Khan Academy has thousands of free videos on a multitude of topics including linear algebra:

  http://www.khanacademy.org/

- The Linear Algebra toolkit:

  http://www.math.odu.edu/∼bogacki/lat/

- Carter, Tapia and Papakonstantinou's online linear algebra resource

  http://ceee.rice.edu/Books/LA/index.html

- S.O.S. Mathematics Matrix Algebra primer:

  http://www.sosmath.com/matrix/matrix.html

- The Numerical Methods Guy on Youtube. Lots of worked examples:

http://www.youtube.com/user/numericalmethodsguy

- Interactive Mathematics. Lots of useful math lessons on many topics:

  http://www.intmath.com/

- Stat Trek. A quick matrix tutorial for statistics students:

  http://stattrek.com/matrix-algebra/matrix.aspx

- Wolfram's Mathworld. An online mathematics encyclopædia:

  http://mathworld.wolfram.com/

- Paul Dawkin's online math notes:

  http://tutorial.math.lamar.edu/

- Math Doctor Bob:

  http://www.youtube.com/user/MathDoctorBob?feature=watch

- Some pictures of how to rotate objects with matrices:

  http://people.cornellcollege.edu/dsherman/visualize-matrix.html

- xkcd. Geek jokes:

  http://xkcd.com/184/

- See the bridge actually fall down:

  http://anothermathgeek.hubpages.com/hub/What-the-Heck-are-Eigenvalues-and-Eigenvectors

D

Sample First Midterm

Here are some worked problems typical for what you might expect on a first midterm examination.

1. Solve the following linear system. Write the solution set in vector form. Check your solution. Write one particular solution and one homogeneous solution, if they exist. What does the solution set look like geometrically?

$$\begin{array}{rcrcrl} x & + & 3y & & & = 4 \\ x & - & 2y & + & z & = 1 \\ 2x & + & y & + & z & = 5 \end{array}$$

2. Consider the system of equations

$$\left\{ \begin{array}{rcrcrcrcr} x & & & - & z & + & 2w & = & -1 \\ x & + & y & + & z & - & w & = & 2 \\ & - & y & - & 2z & + & 3w & = & -3 \\ 5x & + & 2y & - & z & + & 4w & = & 1 \end{array} \right.$$

   (a) Write an augmented matrix for this system.

   (b) Use elementary row operations to find its reduced row echelon form.

   (c) Write the solution set for the system in the form

$$S = \{X_0 + \sum_i \mu_i Y_i : \mu_i \in \mathbb{R}\}.$$

(d) What are the vectors X_0 and Y_i called *and* which matrix equations do they solve?

(e) Check separately that X_0 and each Y_i solve the matrix systems you claimed they solved in part (d).

3. Use row operations to invert the matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 11 \\ 3 & 7 & 14 & 25 \\ 4 & 11 & 25 & 50 \end{pmatrix}$$

4. Let $M = \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix}$. Calculate $M^T M^{-1}$. Is M symmetric? What is the trace of the transpose of $f(M)$, where $f(x) = x^2 - 1$?

5. In this problem M is the matrix

$$M = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

and X is the vector

$$X = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Calculate all possible dot products between the vectors X and MX. Compute the lengths of X and MX. What is the angle between the vectors MX and X. Draw a picture of these vectors in the plane. For what values of θ do you expect equality in the triangle and Cauchy–Schwartz inequalities?

6. Let M be the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Find a formula for M^k for any positive integer power k. Try some simple examples like $k = 2, 3$ if confused.

7. What does it mean for a function to be linear? Check that integration is a linear function from V to V, where $V = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is integrable}\}$ is a vector space over $\mathbb{R}$ with usual addition and scalar multiplication.

8. What are the four main things we need to define for a vector space? Which of the following is a vector space over $\mathbb{R}$? For those that are not vector spaces, modify one part of the definition to make it into a vector space.

   (a) $V = \{\, 2\times 2$ matrices with entries in $\mathbb{R}\}$, usual matrix addition, and $k\cdot\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ka & b \\ kc & d \end{pmatrix}$ for $k\in\mathbb{R}$.

   (b) $V = \{$polynomials with complex coefficients of degree $\leq 3\}$, with usual addition and scalar multiplication of polynomials.

   (c) $V = \{$vectors in $\mathbb{R}^3$ with at least one entry containing a $1\}$, with usual addition and scalar multiplication.

9. *Subspaces:* If V is a vector space, we say that U is a *subspace* of V when the set U is also a vector space, using the vector addition and scalar multiplication rules of the vector space V. (Remember that $U\subset V$ says that "U is a subset of V", *i.e.*, all elements of U are also elements of V. The symbol $\forall$ means "for all" and $\in$ means "is an element of".)

   Explain why additive closure ($u+w\in U\ \forall\ u,v\in U$) and multiplicative closure ($r.u\in U\ \forall\ r\in\mathbb{R},\ u\in V$) ensure that (i) the zero vector $0\in U$ and (ii) every $u\in U$ has an additive inverse.

   In fact it suffices to check closure under addition and scalar multiplication to verify that U is a vector space. Check whether the following choices of U are vector spaces:

   (a) $U = \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} : x,y\in\mathbb{R} \right\}$

   (b) $U = \left\{ \begin{pmatrix} 1 \\ 0 \\ z \end{pmatrix} : z\in\mathbb{R} \right\}$

10. Find an LU decomposition for the matrix

$$\begin{pmatrix} 1 & 1 & -1 & 2 \\ 1 & 3 & 2 & 2 \\ -1 & -3 & -4 & 6 \\ 0 & 4 & 7 & -2 \end{pmatrix}$$

Use your result to solve the system

$$\left\{\begin{array}{rcrcrcrl} x & + & y & - & z & + & 2w & = 7 \\ x & + & 3y & + & 2z & + & 2w & = 6 \\ -x & - & 3y & - & 4z & + & 6w & = 12 \\ & & 4y & + & 7z & - & 2w & = -7 \end{array}\right.$$

Solutions

1. *As an additional exercise, write out the row operations above the $\sim$ signs below.*

$$\left(\begin{array}{rrr|r} 1 & 3 & 0 & 4 \\ 1 & -2 & 1 & 1 \\ 2 & 1 & 1 & 5 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1 & 3 & 0 & 4 \\ 0 & -5 & 1 & -3 \\ 0 & -5 & 1 & -3 \end{array}\right) \sim \left(\begin{array}{rrr|r} 1 & 0 & \frac{3}{5} & \frac{11}{5} \\ 0 & 1 & -\frac{1}{5} & \frac{3}{5} \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Solution set is

$$\left\{\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{11}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix} + \mu \begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix} : \mu \in \mathbb{R}\right\}.$$

Geometrically this represents a line in $\mathbb{R}^3$ through the point $\begin{pmatrix} \frac{11}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix}$ running parallel to the vector $\begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix}$.

The vector $\begin{pmatrix} \frac{11}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix}$ is *a* particular solution and $\begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix}$ is *a* homogeneous solution.

As a double check note that

$$\begin{pmatrix} 1 & 3 & 0 \\ 1 & -2 & 1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{11}{5} \\ \frac{3}{5} \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 5 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 3 & 0 \\ 1 & -2 & 1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} -\frac{3}{5} \\ \frac{1}{5} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

2. (a) The augmented matrix

$$\left(\begin{array}{rrrr|r} 1 & 0 & -1 & 2 & -1 \\ 1 & 1 & 1 & -1 & 2 \\ 0 & -1 & -2 & 3 & -3 \\ 5 & 2 & -1 & 4 & 1 \end{array}\right)$$

encodes the system of equations.

(b) *Again, write out the row operations as an additional exercise.*
The above augmented matrix is row equivalent to

$$\left(\begin{array}{rrrr|r} 1 & 0 & -1 & 2 & -1 \\ 0 & 1 & 2 & -3 & 3 \\ 0 & -1 & -2 & 3 & -3 \\ 0 & 2 & 4 & -6 & 6 \end{array}\right) \sim \left(\begin{array}{rrrr|r} 1 & 0 & -1 & 2 & -1 \\ 0 & 1 & 2 & -3 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

which is in reduced row echelon form.

(c) Solution set is

$$\left\{ X = \begin{pmatrix} -1 \\ 3 \\ 0 \\ 0 \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + \mu_2 \begin{pmatrix} -2 \\ 3 \\ 0 \\ 1 \end{pmatrix} : \mu_1, \mu_2 \in \mathbb{R} \right\} .$$

(d) The vector $X_0 = \begin{pmatrix} -1 \\ 3 \\ 0 \\ 0 \end{pmatrix}$ is *a* particular solution and the vectors

$$Y_1 = \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} \text{ and } Y_2 = \begin{pmatrix} -2 \\ 3 \\ 0 \\ 1 \end{pmatrix}$$

are homogeneous solutions. They obey

$$MX = V \,, \qquad MY_1 = 0 = MY_2 \,.$$

where

$$M = \begin{pmatrix} 1 & 0 & -1 & 2 \\ 1 & 1 & 1 & -1 \\ 0 & -1 & -2 & 3 \\ 5 & 2 & -1 & 4 \end{pmatrix} \text{ and } V = \begin{pmatrix} -1 \\ 2 \\ -3 \\ 1 \end{pmatrix} .$$

(e) This amounts to explicitly performing the matrix manipulations

$$MX - V,\ MY_1,\ \text{and}\ MY_2$$

to verify that they are all zero vectors.

3. *As usual, be sure to write out the row operations above the ∼'s so your work can be easily checked.*

$$\left(\begin{array}{cccc|cccc} 1 & 2 & 3 & 4 & 1 & 0 & 0 & 0 \\ 2 & 4 & 7 & 11 & 0 & 1 & 0 & 0 \\ 3 & 7 & 14 & 25 & 0 & 0 & 1 & 0 \\ 4 & 11 & 25 & 50 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|cccc} 1 & 2 & 3 & 4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & -2 & 1 & 0 & 0 \\ 0 & 1 & 5 & 13 & -3 & 0 & 1 & 0 \\ 0 & 3 & 13 & 34 & -4 & 0 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|cccc} 1 & 0 & -7 & -22 & 7 & 0 & -2 & 0 \\ 0 & 1 & 5 & 13 & -3 & 0 & 1 & 0 \\ 0 & 0 & 1 & 3 & -2 & 1 & 0 & 0 \\ 0 & 0 & -2 & -5 & 5 & 0 & -3 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & -1 & -7 & 7 & -2 & 0 \\ 0 & 1 & 0 & -2 & 7 & -5 & 1 & 0 \\ 0 & 0 & 1 & 3 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 2 & -3 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -6 & 9 & -5 & 1 \\ 0 & 1 & 0 & 0 & 9 & -1 & -5 & 2 \\ 0 & 0 & 1 & 0 & -5 & -5 & 9 & -3 \\ 0 & 0 & 0 & 1 & 1 & 2 & -3 & 1 \end{array}\right).$$

Check

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 11 \\ 3 & 7 & 14 & 25 \\ 4 & 11 & 25 & 50 \end{pmatrix} \begin{pmatrix} -6 & 9 & -5 & 1 \\ 9 & -1 & -5 & 2 \\ -5 & -5 & 9 & -3 \\ 1 & 2 & -3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

4.

$$M^T M^{-1} = \begin{pmatrix} 2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{3}{5} & -\frac{2}{5} \end{pmatrix} = \begin{pmatrix} \frac{11}{5} & -\frac{4}{5} \\ -\frac{2}{5} & \frac{3}{5} \end{pmatrix}.$$

Since $M^T M^{-1} \neq I$, it follows $M^T \neq M$ so M is *not* symmetric. Finally

$$\operatorname{tr} f(M)^T = \operatorname{tr} f(M) = \operatorname{tr}(M^2 - I) = \operatorname{tr} \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix} - \operatorname{tr} I$$

$$= (2 \cdot 2 + 1 \cdot 3) + (3 \cdot 1 + (-1) \cdot (-1)) - 2 = 9\,.$$

5. First

$$X \cdot (MX) = X^T M X = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

$$= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta \end{pmatrix} = (x^2 + y^2)\cos\theta\,.$$

Now $||X|| = \sqrt{X \cdot X} = \sqrt{x^2 + y^2}$ and $(MX) \cdot (MX) = X M^T M X$. But

$$M^T M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{pmatrix} = I\,.$$

Hence $||MX|| = ||X|| = \sqrt{x^2 + y^2}$. Thus the cosine of the angle between X and MX is given by

$$\frac{X \cdot (MX)}{||X||\ ||MX||} = \frac{(x^2 + y^2)\cos\theta}{\sqrt{x^2 + y^2}\,\sqrt{x^2 + y^2}} = \cos\theta\,.$$

In other words, the angle is θ OR $-\theta$. You should draw two pictures, one where the angle between X and MX is θ, the other where it is $-\theta$.

For Cauchy–Schwartz, $\frac{|X \cdot (MX)|}{||X||\ ||MX||} = |\cos\theta| = 1$ when $\theta = 0, \pi$. For the triangle equality $MX = X$ achieves $||X + MX|| = ||X|| + ||MX||$, which requires $\theta = 0$.

6. This is a block matrix problem. Notice the that matrix M is really just $M = \begin{pmatrix} I & I \\ 0 & I \end{pmatrix}$, where I and 0 are the 3×3 identity zero matrices, respectively. But

$$M^2 = \begin{pmatrix} I & I \\ 0 & I \end{pmatrix} \begin{pmatrix} I & I \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & 2I \\ 0 & I \end{pmatrix}$$

and

$$M^3 = \begin{pmatrix} I & I \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 2I \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & 3I \\ 0 & I \end{pmatrix}$$

so, $M^k = \begin{pmatrix} I & kI \\ 0 & I \end{pmatrix}$, or explicitly

$$M^k = \begin{pmatrix} 1 & 0 & 0 & k & 0 & 0 \\ 0 & 1 & 0 & 0 & k & 0 \\ 0 & 0 & 1 & 0 & 0 & k \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

7. We can call a function $f\colon V \longrightarrow W$ *linear* if the sets V and W are vector spaces and f obeys

$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v)\,,$$

for all $u, v \in V$ and $\alpha, \beta \in \mathbb{R}$.

Now, integration is a linear transformation from the space V of all integrable functions (don't be confused between the definition of a linear function above, and integrable functions $f(x)$ which here are the vectors in V) to the real numbers $\mathbb{R}$, because $\int_{-\infty}^{\infty}(\alpha f(x) + \beta g(x))dx = \alpha \int_{-\infty}^{\infty} f(x)dx + \beta \int_{-\infty}^{\infty} g(x)dx$.

8. The four main ingredients are (i) a set V of vectors, (ii) a number field K (usually $K = \mathbb{R}$), (iii) a rule for adding vectors (vector addition) and (iv) a way to multiply vectors by a number to produce a new vector (scalar multiplication). There are, of course, ten rules that these four ingredients must obey.

   (a) This is not a vector space. Notice that distributivity of scalar multiplication requires $2u = (1 + 1)u = u + u$ for any vector u but

$$2 \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2a & b \\ 2c & d \end{pmatrix}$$

which does *not* equal

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2a & 2b \\ 2c & 2d \end{pmatrix}.$$

This could be repaired by taking

$$k \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ka & kb \\ kc & kd \end{pmatrix}.$$

(b) This is a vector space. *Although, the question does not ask you to, it is a useful exercise to verify that all ten vector space rules are satisfied.*

(c) This is not a vector space for many reasons. An easy one is that $(1,-1,0)$ and $(-1,1,0)$ are both in the space, but their sum $(0,0,0)$ is not (*i.e.*, additive closure fails). The easiest way to repair this would be to drop the requirement that there be at least one entry equaling 1.

9. (i) Thanks to multiplicative closure, if $u \in U$, so is $(-1)\cdot u$. But $(-1)\cdot u + u = (-1)\cdot u + 1\cdot u = (-1+1)\cdot u = 0.u = 0$ (at each step in this chain of equalities we have used the fact that V is a vector space and therefore can use its vector space rules). In particular, this means that the zero vector of V is in U and is its zero vector also. (ii) Also, in V, for each u there is an element $-u$ such that $u + (-u) = 0$. But by additive close, $(-u)$ must also be in U, thus every $u \in U$ has an additive inverse.

(a) This is a vector space. First we check additive closure: let $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$ and $\begin{pmatrix} z \\ w \\ 0 \end{pmatrix}$ be arbitrary vectors in U. But since $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} + \begin{pmatrix} z \\ w \\ 0 \end{pmatrix} = \begin{pmatrix} x+z \\ y+w \\ 0 \end{pmatrix}$, so is their sum (because vectors in U are those whose third component vanishes). Multiplicative closure is similar: for any $\alpha \in \mathbb{R}$, $\alpha \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} \alpha x \\ \alpha y \\ 0 \end{pmatrix}$, which also has no third component, so is in U.

(b) This is not a vector space for various reasons. A simple one is that $u = \begin{pmatrix} 1 \\ 0 \\ z \end{pmatrix}$ is in U but the vector $u + u = \begin{pmatrix} 2 \\ 0 \\ 2z \end{pmatrix}$ is not in U (it has a 2 in the first component, but vectors in U always have a 1 there).

10.

$$\begin{pmatrix} 1 & 1 & -1 & 2 \\ 1 & 3 & 2 & 2 \\ -1 & -3 & -4 & 6 \\ 0 & 4 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 2 & 3 & 0 \\ 0 & -2 & -5 & 8 \\ 0 & 4 & 7 & -2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 2 & 3 & 0 \\ 0 & 0 & -2 & 8 \\ 0 & 0 & 1 & -2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & -1 & 1 & 0 \\ 0 & 2 & -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 2 & 3 & 0 \\ 0 & 0 & -2 & 8 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

To solve $MX = V$ using $M = LU$ we first solve $LW = V$ whose augmented matrix reads

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 7 \\ 1 & 1 & 0 & 0 & 6 \\ -1 & -1 & 1 & 0 & 12 \\ 0 & 2 & -\frac{1}{2} & 1 & -7 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 7 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 18 \\ 0 & 2 & -\frac{1}{2} & 1 & -7 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 7 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 18 \\ 0 & 0 & 0 & 1 & 4 \end{array}\right),$$

from which we can read off W. Now we compute X by solving $UX = W$ with the augmented matrix

$$\left(\begin{array}{cccc|c} 1 & 1 & -1 & 2 & 7 \\ 0 & 2 & 3 & 0 & -1 \\ 0 & 0 & -2 & 8 & 18 \\ 0 & 0 & 0 & 2 & 4 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 1 & -1 & 2 & 7 \\ 0 & 2 & 3 & 0 & -1 \\ 0 & 0 & -2 & 0 & 2 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|c} 1 & 1 & -1 & 2 & 7 \\ 0 & 2 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right) \sim \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right).$$

So $x = 1$, $y = 1$, $z = -1$ and $w = 2$.

$\mathcal{E}$

Sample Second Midterm

Here are some worked problems typical for what you might expect on a second midterm examination.

1. *Determinants:* The determinant $\det M$ of a 2×2 matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is defined by
$$\det M = ad - bc\,.$$
   (a) For which values of $\det M$ does M have an inverse?
   (b) Write down all 2×2 bit matrices with determinant 1. (Remember bits are either 0 or 1 and $1 + 1 = 0$.)
   (c) Write down all 2×2 bit matrices with determinant 0.
   (d) Use one of the above examples to show why the following statement is FALSE.

   Square matrices with the same determinant are always row equivalent.

2. Let
$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$
Compute $\det A$. Find all solutions to (i) $AX = 0$ and (ii) $AX = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ for

the vector $X \in \mathbb{R}^3$. Find, but do not solve, the characteristic polynomial of A.

3. Let M be any 2×2 matrix. Show

$$\det M = -\frac{1}{2}\mathrm{tr}M^2 + \frac{1}{2}(\mathrm{tr}M)^2\,.$$

4. *The permanent:* Let $M = (M^i_j)$ be an $n \times n$ matrix. An operation producing a single number from M similar to the determinant is the "permanent"

$$\mathrm{perm}\, M = \sum_\sigma M^1_{\sigma(1)} M^2_{\sigma(2)} \cdots M^n_{\sigma(n)}\,.$$

For example

$$\mathrm{perm}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad + bc\,.$$

Calculate

$$\mathrm{perm}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\,.$$

What do you think would happen to the permanent of an $n \times n$ matrix M if (include a *brief* explanation with each answer):

(a) You multiplied M by a number λ.

(b) You multiplied a row of M by a number λ.

(c) You took the transpose of M.

(d) You swapped two rows of M.

5. Let X be an $n \times 1$ matrix subject to

$$X^T X = (1)\,,$$

and define

$$H = I - 2XX^T\,,$$

(where I is the $n \times n$ identity matrix). Show

$$H = H^T = H^{-1}.$$

6. Suppose λ is an eigenvalue of the matrix M with associated eigenvector v. Is v an eigenvector of M^k (where k is any positive integer)? If so, what would the associated eigenvalue be?

   Now suppose that the matrix N is *nilpotent*, *i.e.*

   $$N^k = 0$$

   for some integer $k \geq 2$. Show that 0 is the only eigenvalue of N.

7. Let $M = \begin{pmatrix} 3 & -5 \\ 1 & -3 \end{pmatrix}$. Compute M^{12}. (Hint: $2^{12} = 4096$.)

8. *The Cayley Hamilton Theorem*: Calculate the characteristic polynomial $P_M(\lambda)$ of the matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Now compute the matrix polynomial $P_M(M)$. What do you observe? Now suppose the $n \times n$ matrix A is "similar" to a diagonal matrix D, in other words

   $$A = P^{-1}DP$$

   for some invertible matrix P and D is a matrix with values $\lambda_1,\ \lambda_2, \ldots \lambda_n$ along its diagonal. Show that the two matrix polynomials $P_A(A)$ and $P_A(D)$ are similar (*i.e.* $P_A(A) = P^{-1}P_A(D)P$). Finally, compute $P_A(D)$, what can you say about $P_A(A)$?

9. *Define* what it means for a set U to be a subspace of a vector space V. Now let U and W be non-trivial subspaces of V. Are the following also subspaces? (Remember that $\cup$ means "union" and $\cap$ means "intersection".)

   (a) $U \cup W$

   (b) $U \cap W$

   In each case *draw* examples in $\mathbb{R}^3$ that justify your answers. If you answered "yes" to either part also give a general explanation why this is the case.

10. *Define* what it means for a set of vectors $\{v_1, v_2, \ldots, v_n\}$ to (i) be linearly independent, (ii) span a vector space V and (iii) be a basis for a vector space V.

    Consider the following vectors in $\mathbb{R}^3$

    $$u = \begin{pmatrix} -1 \\ -4 \\ 3 \end{pmatrix}, \qquad v = \begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix}, \qquad w = \begin{pmatrix} 10 \\ 7 \\ h+3 \end{pmatrix}.$$

    For which values of h is $\{u, v, w\}$ a basis for $\mathbb{R}^3$?

Solutions

1. (a) Whenever $\det M = ad - bc \neq 0$.

   (b) Unit determinant bit matrices:

$$\begin{pmatrix}1&0\\0&1\end{pmatrix},\begin{pmatrix}1&1\\0&1\end{pmatrix},\begin{pmatrix}1&0\\1&1\end{pmatrix},\begin{pmatrix}0&1\\1&0\end{pmatrix},\begin{pmatrix}1&1\\1&0\end{pmatrix},\begin{pmatrix}0&1\\1&1\end{pmatrix}.$$

   (c) Bit matrices with vanishing determinant:

$$\begin{pmatrix}0&0\\0&0\end{pmatrix},\begin{pmatrix}1&0\\0&0\end{pmatrix},\begin{pmatrix}0&1\\0&0\end{pmatrix},\begin{pmatrix}0&0\\1&0\end{pmatrix},\begin{pmatrix}0&0\\0&1\end{pmatrix},$$

$$\begin{pmatrix}1&1\\0&0\end{pmatrix},\begin{pmatrix}0&0\\1&1\end{pmatrix},\begin{pmatrix}1&0\\1&0\end{pmatrix},\begin{pmatrix}0&1\\0&1\end{pmatrix},\begin{pmatrix}1&1\\1&1\end{pmatrix}.$$

   As a check, count that the total number of 2×2 *bit matrices is* $2^{(\text{number of entries})} = 2^4 = 16$.

   (d) To disprove this statement, we just need to find a single counterexample. All the unit determinant examples above are actually row equivalent to the identity matrix, so focus on the bit matrices with vanishing determinant. Then notice (for example), that

$$\begin{pmatrix}1&1\\0&0\end{pmatrix}\not\sim\begin{pmatrix}0&0\\0&0\end{pmatrix}.$$

   So we have found a pair of matrices that are not row equivalent but do have the same determinant. It follows that the statement is false.

2.

$$\det A = 1.(2.6-3.5) - 1.(2.6-3.4) + 1.(2.5-2.4) = -1\,.$$

(i) Since $\det A \neq 0$, the homogeneous system $AX = 0$ only has the solution $X = 0$. (ii) It is efficient to compute the adjoint

$$\text{adj } A = \begin{pmatrix}-3&0&2\\-1&2&-1\\1&-1&0\end{pmatrix}^T = \begin{pmatrix}-3&-1&1\\0&2&-1\\2&-1&0\end{pmatrix}$$

Hence

$$A^{-1} = \begin{pmatrix}3&1&-1\\0&-2&1\\-2&1&0\end{pmatrix}.$$

Thus

$$X = \begin{pmatrix} 3 & 1 & -1 \\ 0 & -2 & 1 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} .$$

Finally,

$$P_A(\lambda) = -\det \begin{pmatrix} 1-\lambda & 1 & 1 \\ 2 & 2-\lambda & 3 \\ 4 & 5 & 6-\lambda \end{pmatrix}$$

$$= -\Big[(1-\lambda)[(2-\lambda)(6-\lambda)-15] - [2.(6-\lambda)-12] + [10-4.(2-\lambda)]\Big]$$

$$= \lambda^3 - 9\lambda^2 - \lambda + 1 .$$

3. Call $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then $\det M = ad - bc$, yet

$$-\frac{1}{2}\operatorname{tr} M^2 + \frac{1}{2}(\operatorname{tr} M)^2 = -\frac{1}{2}\operatorname{tr}\begin{pmatrix} a^2+bc & * \\ * & bc+d^2 \end{pmatrix} - \frac{1}{2}(a+d)^2$$

$$= -\frac{1}{2}(a^2+2bc+d^2) + \frac{1}{2}(a^2+2ad+d^2) = ad - bc ,$$

which is what we were asked to show.

4.

$$\operatorname{perm}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = 1\cdot(5\cdot 9 + 6\cdot 8) + 2\cdot(4\cdot 9 + 6\cdot 7) + 3\cdot(4\cdot 8 + 5\cdot 7) = 450 .$$

(a) Multiplying M by λ replaces every matrix element $M^i_{\sigma(j)}$ in the formula for the permanent by $\lambda M^i_{\sigma(j)}$, and therefore produces an overall factor λ^n.

(b) Multiplying the ith row by λ replaces $M^i_{\sigma(j)}$ in the formula for the permanent by $\lambda M^i_{\sigma(j)}$. Therefore the permanent is multiplied by an overall factor λ.

(c) The permanent of a matrix transposed equals the permanent of the original matrix, because in the formula for the permanent this amounts to summing over permutations of rows rather than columns. But we could then sort the product $M_1^{\sigma(1)} M_2^{\sigma(2)} \ldots M_n^{\sigma(n)}$ back into its original order using the inverse permutation σ^{-1}. But summing over permutations is equivalent to summing over inverse permutations, and therefore the permanent is unchanged.

(d) Swapping two rows also leaves the permanent unchanged. The argument is almost the same as in the previous part, except that we need only reshuffle two matrix elements $M^j_{\sigma(i)}$ and $M^i_{\sigma(j)}$ (in the case where rows i and j were swapped). Then we use the fact that summing over all permutations σ or over all permutations $\widetilde{\sigma}$ obtained by swapping a pair in σ are equivalent operations.

5. Firstly, lets call $(1) = 1$ (the 1×1 identity matrix). Then we calculate

$$H^T = (I-2XX^T)^T = I^T - 2(XX^T)^T = I - 2(X^T)^T X^T = I - 2XX^T = H\,,$$

which demonstrates the first equality. Now we compute

$$H^2 = (I-2XX^T)(I-2XX^T) = I - 4XX^T + 4XX^TXX^T$$

$$= I - 4XX^T + 4X(X^TX)X^T = I - 4XX^T + 4X.1.X^T = I\,.$$

So, since $HH = I$, we have $H^{-1} = H$.

6. We know $Mv = \lambda v$. Hence

$$M^2v = MMv = M\lambda v = \lambda Mv = \lambda^2 v\,,$$

and similarly

$$M^k v = \lambda M^{k-1}v = \ldots = \lambda^k v\,.$$

So v is an eigenvector of M^k with eigenvalue λ^k.

Now let us assume v is an eigenvector of the nilpotent matrix N with eigenvalue λ. Then from above

$$N^k v = \lambda^k v$$

but by nilpotence, we also have

$$N^k v = 0.$$

Hence $\lambda^k v = 0$ and v (being an eigenvector) cannot vanish. Thus $\lambda^k = 0$ and in turn $\lambda = 0$.

7. Let us think about the eigenvalue problem $Mv = \lambda v$. This has solutions when

$$0 = \det\begin{pmatrix} 3-\lambda & -5 \\ 1 & -3-\lambda \end{pmatrix} = \lambda^2 - 4 \Rightarrow \lambda = \pm 2\,.$$

The associated eigenvalues solve the homogeneous systems (in augmented matrix form)

$$\left(\begin{array}{cc|c} 1 & -5 & 0 \\ 1 & -5 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -5 & 0 \\ 0 & 0 & 0 \end{array}\right) \text{ and } \left(\begin{array}{cc|c} 5 & -5 & 0 \\ 1 & -1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right),$$

respectively, so are $v_2 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$ and $v_{-2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Hence $M^{12}v_2 = 2^{12}v_2$ and $M^{12}v_{-2} = (-2)^{12}v_{-2}$. Now, $\begin{pmatrix} x \\ y \end{pmatrix} = \frac{x-y}{4}\begin{pmatrix} 5 \\ 1 \end{pmatrix} - \frac{x-5y}{4}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ (this was obtained by solving the linear system $av_2 + bv_{-2} =$ for a and b). Thus

$$M\begin{pmatrix} x \\ y \end{pmatrix} = \frac{x-y}{4}Mv_2 - \frac{x-5y}{4}Mv_{-2}$$

$$= 2^{12}\Big(\frac{x-y}{4}v_2 - \frac{x-5y}{4}v_{-2}\Big) = 2^{12}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Thus

$$M^{12} = \begin{pmatrix} 4096 & 0 \\ 0 & 4096 \end{pmatrix}.$$

If you understand the above explanation, then you have a good understanding of diagonalization. A quicker route is simply to observe that $M^2 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$.

8.

$$P_M(\lambda) = (-1)^2\det\begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} = (\lambda - a)(\lambda - d) - bc\,.$$

Thus

$$P_M(M) = (M - aI)(M - dI) - bcI$$

$$= \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}\right)\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix}\right) - \begin{pmatrix} bc & 0 \\ 0 & bc \end{pmatrix}$$

$$= \begin{pmatrix} 0 & b \\ c & d-a \end{pmatrix}\begin{pmatrix} a-d & b \\ c & 0 \end{pmatrix} - \begin{pmatrix} bc & 0 \\ 0 & bc \end{pmatrix} = 0\,.$$

Observe that any 2×2 matrix is a zero of its own characteristic polynomial (*in fact this holds for square matrices of any size*).

Now if $A = P^{-1}DP$ then $A^2 = P^{-1}DPP^{-1}DP = P^{-1}D^2P$. Similarly $A^k = P^{-1}D^kP$. So for *any* matrix polynomial we have

$$\begin{aligned} & A^n + c_1A^{n-1} + \cdots c_{n-1}A + c_nI \\ = \;& P^{-1}D^nP + c_1P^{-1}D^{n-1}P + \cdots c_{n-1}P^{-1}DP + c_nP^{-1}P \\ = \;& P^{-1}(D^n + c_1D^{n-1} + \cdots c_{n-1}D + c_nI)P\,. \end{aligned}$$

Thus we may conclude $P_A(A) = P^{-1}P_A(D)P$.

Now suppose $D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & \lambda_n \end{pmatrix}$. Then

$$P_A(\lambda) = \det(\lambda I - A) = \det(\lambda P^{-1}IP - P^{-1}DP) = \det P.\det(\lambda I - D).\det P$$

$$= \det(\lambda I - D) = \det \begin{pmatrix} \lambda - \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda - \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda - \lambda_n \end{pmatrix}$$

$$= (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)\,.$$

Thus we see that $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of M. Finally we compute

$$P_A(D) = (D - \lambda_1)(D - \lambda_2) \dots (D - \lambda_n)$$

$$= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & 0 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \cdots \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} = 0\,.$$

We conclude the $P_M(M) = 0$.

9. A subset of a vector space is called a subspace if it itself is a vector space, using the rules for vector addition and scalar multiplication inherited from the original vector space.

   (a) So long as $U \neq U \cup W \neq W$ the answer is *no*. Take, for example, U to be the x-axis in $\mathbb{R}^2$ and W to be the y-axis. Then $\big(1, 0\big) \in U$ and $\big(0, 1\big) \in W$, but $\big(1, 0\big) + \big(0, 1\big) = \big(1, 1\big) \notin U \cup W$. So $U \cup W$ is not additively closed and is not a vector space (and thus not a subspace). It is easy to draw the example described.

   (b) Here the answer is always *yes*. The proof is not difficult. Take a vector u and w such that $u \in U \cap W \ni w$. This means that *both* u and w are in *both* U and W. But, since U is a vector space, $\alpha u + \beta w$ is also in U. Similarly, $\alpha u + \beta w \in W$. Hence $\alpha u + \beta w \in U \cap W$. So closure holds in $U \cap W$ and this set is a subspace by the subspace theorem. Here, a good picture to draw is two planes through the origin in $\mathbb{R}^3$ intersecting at a line (also through the origin).

10. (i) We say that the vectors $\{v_1, v_2, \dots v_n\}$ are linearly independent if there exist *no* constants c^1, $c^2, \dots c^n$ (not all vanishing) such that $c^1v_1 + c^2v_2 + \cdots + c^nv_n = 0$. Alternatively, we can require that there is no non-trivial solution for scalars c^1, $c^2, \dots, c^n$ to the linear system $c^1v_1 + c^2v_2 + \cdots + c^nv_n = 0$. (ii) We say that these vectors span a vector space V if the set $\text{span}\{v_1, v_2, \dots v_n\} = \{c^1v_1 + c^2v_2 + \cdots + c^nv_n : c^1, c^2, \dots c^n \in \mathbb{R}\} = V$. (iii) We call $\{v_1, v_2, \dots v_n\}$ a basis for V if $\{v_1, v_2, \dots v_n\}$ are linearly independent *and* $\text{span}\{v_1, v_2, \dots v_n\} = V$.

    For u, v, w to be a basis for $\mathbb{R}^3$, we firstly need (the spanning requirement) that any vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ can be written as a linear combination of u, v and w

$$c^1 \begin{pmatrix} -1 \\ -4 \\ 3 \end{pmatrix} + c^2 \begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix} + c^3 \begin{pmatrix} 10 \\ 7 \\ h+3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

    The linear independence requirement implies that when $x = y = z = 0$, the only solution to the above system is $c^1 = c^2 = c^3 = 0$. But the above system in matrix language reads

$$\begin{pmatrix} -1 & 4 & 10 \\ -4 & 5 & 7 \\ 3 & 0 & h+3 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \\ c^3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

    Both requirements mean that the matrix on the left hand side must be invertible, so we examine its determinant

$$\det \begin{pmatrix} -1 & 4 & 10 \\ -4 & 5 & 7 \\ 3 & 0 & h+3 \end{pmatrix} = -4 \cdot (-4 \cdot (h+3) - 7 \cdot 3) + 5 \cdot (-1 \cdot (h+3) - 10 \cdot 3)$$

$$= 11(h-3) \cdot$$

    Hence we obtain a basis whenever $h \neq 3$.

F

Sample Final Exam

Here are some worked problems typical for what you might expect on a final examination.

1. Define the following terms:

   (a) An *orthogonal matrix*.

   (b) A *basis* for a vector space.

   (c) The *span* of a set of vectors.

   (d) The *dimension* of a vector space.

   (e) An *eigenvector*.

   (f) A *subspace* of a vector space.

   (g) The *kernel* of a linear transformation.

   (h) The *nullity* of a linear transformation.

   (i) The *image* of a linear transformation.

   (j) The *rank* of a linear transformation.

   (k) The *characteristic polynomial* of a square matrix.

   (l) An *equivalence relation*.

   (m) A *homogeneous solution* to a linear system of equations.

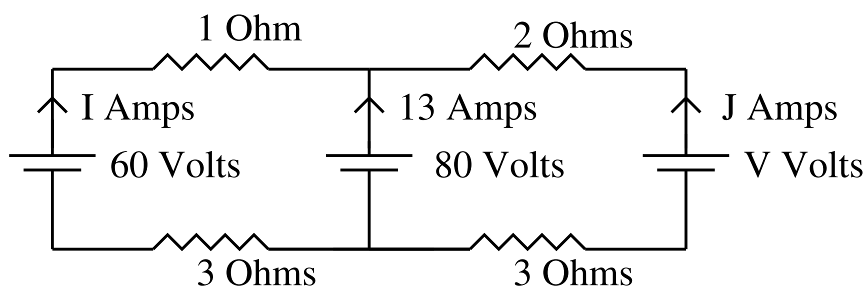
   (n) A *particular solution* to a linear system of equations.

   (o) The *general solution* to a linear system of equations.

   (p) The *direct sum* of a pair of subspaces of a vector space.

(q) The *orthogonal complement* to a subspace of a vector space.

2. *Kirchoff's laws*: Electrical circuits are easy to analyze using systems of equations. The change in voltage (measured in Volts) around any loop due to batteries || and resistors /\/\/\/\ (given by the product of the current measured in Amps and resistance measured in Ohms) equals zero. Also, the sum of currents entering any junction vanishes. Consider the circuit



Find all possible equations for the unknowns I, J and V and then solve for I, J and V. Give your answers with correct units.

3. Suppose M is the matrix of a linear transformation

$$L : U \to V$$

and the vector spaces U and V have dimensions

$$\dim U = n\,, \qquad \dim V = m\,,$$

and

$$m \neq n\,.$$

Also assume

$$\ker L = \{0_U\}\,.$$

(a) How many rows does M have?

(b) How many columns does M have?

(c) Are the columns of M linearly independent?

(d) What size matrix is $M^T M$?

(e) What size matrix is $M M^T$?

(f) Is $M^T M$ invertible?

(g) is $M^T M$ symmetric?

(h) Is $M^T M$ diagonalizable?

(i) Does $M^T M$ have a zero eigenvalue?

(j) Suppose $U = V$ and $\ker L \neq \{0_U\}$. Find an eigenvalue of M.

(k) Suppose $U = V$ and $\ker L \neq \{0_U\}$. Find $\det M$.

4. Consider the system of equations

$$\begin{array}{ccccccccc} x & + & y & + & z & + & w & = & 1 \\ x & + & 2y & + & 2z & + & 2w & = & 1 \\ x & + & 2y & + & 3z & + & 3w & = & 1 \end{array}$$

Express this system as a matrix equation $MX = V$ and then find the solution set by computing an LU decomposition for the matrix M (be sure to use back and forward substitution).

5. Compute the following determinants

$$\det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \ \det\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \ \det\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{pmatrix},$$

$$\det\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 & 25 \end{pmatrix}.$$

Now test your skills on

$$\det\begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ n+1 & n+2 & n+3 & \cdots & 2n \\ 2n+1 & 2n+2 & 2n+3 & & 3n \\ \vdots & & & \ddots & \vdots \\ n^2-n+1 & n^2-n+2 & n^2-n+3 & \cdots & n^2 \end{pmatrix}.$$

Make sure to jot down a few brief notes explaining any clever tricks you use.

6. For which values of a does

$$U = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}, \begin{pmatrix} a \\ 1 \\ 0 \end{pmatrix} \right\} = \mathbb{R}^3 \ ?$$

For any special values of a at which $U \neq \mathbb{R}^3$, express the subspace U as the span of the least number of vectors possible. Give the dimension of U for these cases and draw a picture showing U inside $\mathbb{R}^3$.

7. *Vandermonde determinant:* Calculate the following determinants

$$\det\begin{pmatrix}1 & x\\ 1 & y\end{pmatrix}, \quad \det\begin{pmatrix}1 & x & x^2\\ 1 & y & y^2\\ 1 & z & z^2\end{pmatrix}, \quad \det\begin{pmatrix}1 & x & x^2 & x^3\\ 1 & y & y^2 & y^3\\ 1 & z & z^2 & z^3\\ 1 & w & w^2 & w^3\end{pmatrix}.$$

Be sure to factorize you answers, if possible.

Challenging: Compute the determinant

$$\det\begin{pmatrix}1 & x_1 & (x_1)^2 & \cdots & (x_1)^{n-1}\\ 1 & x_2 & (x_2)^2 & \cdots & (x_2)^{n-1}\\ 1 & x_3 & (x_3)^2 & \cdots & (x_3)^{n-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_n & (x_n)^2 & \cdots & (x_n)^{n-1}\end{pmatrix}.$$

8. (a) Do the vectors $\left\{\begin{pmatrix}1\\2\\3\end{pmatrix}, \begin{pmatrix}3\\2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$ form a basis for $\mathbb{R}^3$? *Be sure to justify your answer.*

   (b) Find a basis for $\mathbb{R}^4$ that includes the vectors $\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$ and $\begin{pmatrix}4\\3\\2\\1\end{pmatrix}$.

   (c) Explain in words how to generalize your computation in part (b) to obtain a basis for $\mathbb{R}^n$ that includes a given pair of (linearly independent) vectors u and v.

9. Elite NASA engineers determine that if a satellite is placed in orbit starting at a point $\mathcal{O}$, it will return exactly to that same point after one orbit of the earth. Unfortunately, if there is a small mistake in the original location of the satellite, which the engineers label by a vector X in $\mathbb{R}^3$ with origin[1] at $\mathcal{O}$,

[1] This is a spy satellite. The exact location of $\mathcal{O}$, the orientation of the coordinate axes in $\mathbb{R}^3$ and the unit system employed by the engineers are CIA secrets.

after one orbit the satellite will instead return to some other point $Y \in \mathbb{R}^3$. The engineer's computations show that Y is related to X by a matrix

$$Y = \begin{pmatrix} 0 & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 \end{pmatrix} X \, .$$

(a) Find all eigenvalues of the above matrix.

(b) Determine *all* possible eigenvectors associated with each eigenvalue.

Let us assume that the rule found by the engineers applies to all subsequent orbits. Discuss case by case, what will happen to the satellite if the initial mistake in its location is in a direction given by an eigenvector.

10. In this problem the scalars in the vector spaces are bits ($0, 1$ with $1+1=0$). The space B^k is the vector space of bit-valued, k-component column vectors.

(a) Find a basis for B^3.

(b) Your answer to part (a) should be a list of vectors v_1, $v_2, \ldots v_n$. What number did you find for n?

(c) How many elements are there in the *set* B^3.

(d) What is the dimension of the vector space B^3.

(e) Suppose $L : B^3 \to B = \{0, 1\}$ is a linear transformation. Explain why specifying $L(v_1)$, $L(v_2), \ldots, L(v_n)$ completely determines L.

(f) Use the notation of part (e) to list all linear transformations

$$L : B^3 \to B \, .$$

How many different linear transformations did you find? Compare your answer to part (c).

(g) Suppose $L_1 : B^3 \to B$ and $L_2 : B^3 \to B$ are linear transformations, and α and β are bits. Define a new map $(\alpha L_1 + \beta L_2) : B^3 \to B$ by

$$(\alpha L_1 + \beta L_2)(v) = \alpha L_1(v) + \beta L_2(v).$$

Is this map a linear transformation? Explain.

(h) Do you think the set of all linear transformations from B^3 to B is a vector space using the addition rule above? If you answer yes, give a basis for this vector space and state its dimension.

11. A team of distinguished, post-doctoral engineers analyzes the design for a bridge across the English channel. They notice that the force on the center of the bridge when it is displaced by an amount $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is given by

$$F = \begin{pmatrix} -x-y \\ -x-2y-z \\ -y-z \end{pmatrix}.$$

Moreover, having read Newton's Principiæ, they know that force is proportional to acceleration so that[2]

$$F = \frac{d^2X}{dt^2}.$$

Since the engineers are worried the bridge might start swaying in the heavy channel winds, they search for an oscillatory solution to this equation of the form[3]

$$X = \cos(\omega t) \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

(a) By plugging their proposed solution in the above equations the engineers find an eigenvalue problem

$$M \begin{pmatrix} a \\ b \\ c \end{pmatrix} = -\omega^2 \begin{pmatrix} a \\ b \\ c \end{pmatrix}.$$

Here M is a 3×3 matrix. Which 3×3 matrix M did the engineers find? *Justify your answer.*

(b) Find the eigenvalues and eigenvectors of the matrix M.

(c) The number $|\omega|$ is often called a *characteristic frequency*. What characteristic frequencies do you find for the proposed bridge?

(d) Find an orthogonal matrix P such that $MP = PD$ where D is a diagonal matrix. *Be sure to also state your result for D.*

---

[2] The bridge is intended for French and English military vehicles, so the exact units, coordinate system and constant of proportionality are state secrets.

[3] Here, a, b, c and ω are constants which we aim to calculate.

(e) Is there a direction in which displacing the bridge yields no force? If so give a vector in that direction. *Briefly* evaluate the quality of this bridge design.

12. *Conic Sections*: The equation for the most general conic section is given by

$$ax^2 + 2bxy + dy^2 + 2cx + 2ey + f = 0\,.$$

Our aim is to analyze the solutions to this equation using matrices.

(a) Rewrite the above quadratic equation as one of the form

$$X^T MX + X^T C + C^T X + f = 0$$

relating an unknown column vector $X = \begin{pmatrix} x \\ y \end{pmatrix}$, its transpose X^T, a 2×2 matrix M, a constant column vector C and the constant f.

(b) Does your matrix M obey any special properties? Find its eigenvalues. You may call your answers λ and μ for the rest of the problem to save writing.

For the rest of this problem we will focus on central conics for which the matrix M is invertible.

(c) Your equation in part (a) above should be be quadratic in X. Recall that if $m \neq 0$, the quadratic equation $mx^2 + 2cx + f = 0$ can be rewritten by *completing the square*

$$m\Big(x + \frac{c}{m}\Big)^2 = \frac{c^2}{m} - f\,.$$

Being very careful that you are now dealing with matrices, use the same trick to rewrite your answer to part (a) in the form

$$Y^T MY = g.$$

Make sure you give formulas for the new unknown column vector Y and constant g in terms of X, M, C and f. You need not multiply out any of the matrix expressions you find.

If all has gone well, you have found a way to shift coordinates for the original conic equation to a new coordinate system with its origin at the center of symmetry. Our next aim is to rotate the coordinate axes to produce a readily recognizable equation.

(d) Why is the angle between vectors V and W is not changed when you replace them by PV and PW for P any orthogonal matrix?

(e) Explain how to choose an orthogonal matrix P such that $MP = PD$ where D is a diagonal matrix.

(f) For the choice of P above, define our final unknown vector Z by $Y = PZ$. Find an expression for $Y^T MY$ in terms of Z and the eigenvalues of M.

(g) Call $Z = \begin{pmatrix} z \\ w \end{pmatrix}$. What equation do z and w obey? (Hint, write your answer using λ, μ and g.)

(h) Central conics are circles, ellipses, hyperbolae or a pair of straight lines. Give examples of values of (λ, μ, g) which produce each of these cases.

13. Let $L\colon V \to W$ be a linear transformation between finite-dimensional vector spaces V and W, and let M be a matrix for L (with respect to some basis for V and some basis for W). We know that L has an inverse if and only if it is bijective, and we know a lot of ways to tell whether M has an inverse. In fact, L has an inverse if and only if M has an inverse:

    (a) Suppose that L is bijective (i.e., one-to-one and onto).

    i. Show that $\dim V = \operatorname{rank} L = \dim W$.

    ii. Show that 0 is not an eigenvalue of M.

    iii. Show that M is an invertible matrix.

    (b) Now, suppose that M is an invertible matrix.

    i. Show that 0 is not an eigenvalue of M.

    ii. Show that L is injective.

    iii. Show that L is surjective.

14. Captain Conundrum gives Queen Quandary a pair of newborn doves, male and female for her birthday. After one year, this pair of doves breed and produce a pair of dove eggs. One year later these eggs hatch yielding a new pair of doves while the original pair of doves breed again and an additional pair of eggs are laid. Captain Conundrum is very happy because now he will never need to buy the Queen a present ever again!

    Let us say that in year zero, the Queen has no doves. In year one she has one pair of doves, in year two she has two pairs of doves *etc...* Call F_n the number of pairs of doves in years n. For example, $F_0 = 0$, $F_1 = 1$ and $F_2 = 1$. Assume no doves die and that the same breeding pattern continues

well into the future. Then $F_3 = 2$ because the eggs laid by the first pair of doves in year two hatch. Notice also that in year three, two pairs of eggs are laid (by the first and second pair of doves). Thus $F_4 = 3$.

(a) Compute F_5 and F_6.

(b) Explain why (for any $n \geq 2$) the following *recursion relation* holds

$$F_n = F_{n-1} + F_{n-2}\,.$$

(c) Let us introduce a column vector $X_n = \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix}$. Compute X_1 and X_2. Verify that these vectors obey the relationship

$$X_2 = MX_1 \text{ where } M = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

(d) Show that $X_{n+1} = MX_n$.

(e) Diagonalize M. (*I.e.*, write M as a product $M = PDP^{-1}$ where D is diagonal.)

(f) Find a simple expression for M^n in terms of P, D and P^{-1}.

(g) Show that $X_{n+1} = M^n X_1$.

(h) The number

$$\varphi = \frac{1+\sqrt{5}}{2}$$

is called the *golden ratio*. Write the eigenvalues of M in terms of φ.

(i) Put your results from parts (c), (f) and (g) together (along with a short matrix computation) to find the formula for the number of doves F_n in year n expressed in terms of φ, $1-\varphi$ and n.

15. Use Gram–Schmidt to find an orthonormal basis for

$$\operatorname{span}\left\{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\0\\1\\1\end{pmatrix}, \begin{pmatrix}0\\0\\1\\2\end{pmatrix}\right\}.$$

16. Let M be the matrix of a linear transformation $L : V \to W$ in given bases for V and W. Fill in the blanks below with one of the following six vector spaces: V, W, $\ker L$, $\big(\ker L\big)^{\perp}$, $\operatorname{im} L$, $\big(\operatorname{im} L\big)^{\perp}$.

(a) The columns of M span ______ in the basis given for ______.

(b) The rows of M span ______ in the basis given for ______.

Suppose

$$M = \begin{pmatrix} 1 & 2 & 1 & 3 \\ 2 & 1 & -1 & 2 \\ 1 & 0 & 0 & -1 \\ 4 & 1 & -1 & 0 \end{pmatrix}$$

is the matrix of L in the bases $\{v_1, v_2, v_3, v_4\}$ for V and $\{w_1, w_2, w_3, w_4\}$ for W. Find bases for ker$L$ and im$L$. Use the dimension formula to check your result.

17. Captain Conundrum collects the following data set

y	x
5	-2
2	-1
0	1
3	2

which he believes to be well-approximated by a parabola

$$y = ax^2 + bx + c\,.$$

(a) Write down a system of four linear equations for the unknown coefficients a, b and c.

(b) Write the augmented matrix for this system of equations.

(c) Find the reduced row echelon form for this augmented matrix.

(d) Are there any solutions to this system?

(e) Find the least squares solution to the system.

(f) What value does Captain Conundrum predict for y when $x = 2$?

18. Suppose you have collected the following data for an experiment

x	y
x_1	y_1
x_2	y_2
x_3	y_3

and believe that the result is well modeled by a straight line

$$y = mx + b\,.$$

(a) Write down a linear system of equations you could use to find the slope m and constant term b.

(b) Arrange the unknowns (m, b) in a column vector X and write your answer to (a) as a matrix equation

$$MX = V\,.$$

Be sure to give explicit expressions for the matrix M and column vector V.

(c) For a generic data set, would you expect your system of equations to have a solution? *Briefly* explain your answer.

(d) Calculate $M^T M$ and $(M^T M)^{-1}$ (for the latter computation, state the condition required for the inverse to exist).

(e) Compute the least squares solution for m and b.

(f) The least squares method determines a vector X that minimizes the length of the vector $V - MX$. Draw a rough sketch of the three data points in the (x, y)-plane as well as their least squares fit. Indicate how the components of $V - MX$ could be obtained from your picture.

Solutions

1. You can find the definitions for all these terms by consulting the index of this book.

2. Both junctions give the same equation for the currents

$$I + J + 13 = 0\,.$$

There are three voltage loops (one on the left, one on the right and one going around the outside of the circuit). Respectively, they give the equations

$$\begin{aligned} 60 - I - 80 - 3I &= 0 \\ 80 + 2J - V + 3J &= 0 \\ 60 - I + 2J - V + 3J - 3I &= 0 \quad . \end{aligned} \tag{F.1}$$

The above equations are easily solved (either using an augmented matrix and row reducing, or by substitution). The result is $I = -5$ Amps, $J = -8$ Amps, $V = 40$ Volts.

3. (a) m.

(b) n.

(c) Yes.

(d) $n \times n$.

(e) $m \times m$.

(f) Yes. This relies on $\ker M = 0$ because if $M^T M$ had a non-trivial kernel, then there would be a non-zero solution X to $M^T M X = 0$. But then by multiplying on the left by X^T we see that $||MX|| = 0$. This in turn implies $MX = 0$ which contradicts the triviality of the kernel of M.

(g) Yes because $\big(M^T M\big)^T = M^T (M^T)^T = M^T M$.

(h) Yes, all symmetric matrices have a basis of eigenvectors.

(i) No, because otherwise it would not be invertible.

(j) Since the kernel of L is non-trivial, M must have 0 as an eigenvalue.

(k) Since M has a zero eigenvalue in this case, its determinant must vanish. I.e., $\det M = 0$.

4. To begin with the system becomes

$$\begin{pmatrix}1 & 1 & 1 & 1\\ 1 & 2 & 2 & 2\\ 1 & 2 & 3 & 3\end{pmatrix}\begin{pmatrix}x\\ y\\ z\\ w\end{pmatrix} = \begin{pmatrix}1\\ 1\\ 1\end{pmatrix}$$

Then

$$M = \begin{pmatrix}1 & 1 & 1 & 1\\ 1 & 2 & 2 & 2\\ 1 & 2 & 3 & 3\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ 1 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 1 & 1 & 1\\ 0 & 1 & 1 & 1\\ 0 & 1 & 2 & 2\end{pmatrix}$$

$$= \begin{pmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}1 & 1 & 1 & 1\\ 0 & 1 & 1 & 1\\ 0 & 0 & 1 & 1\end{pmatrix} = LU$$

So now $MX = V$ becomes $LW = V$ where $W = UX = \begin{pmatrix}a\\ b\\ c\end{pmatrix}$ (say). Thus we solve $LW = V$ by forward substitution

$$a = 1,\ a + b = 1,\ a + b + c = 1 \Rightarrow a = 1, b = 0, c = 0\,.$$

Now solve $UX = W$ by back substitution

$$x + y + z + w = 1,\ y + z + w = 0,\ z + w = 0$$

$$\Rightarrow w = \mu \text{ (arbitrary)}, z = -\mu, y = 0, x = 1\,.$$

The solution set is $\left\{ \begin{pmatrix} x \\ y \\ z \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -\mu \\ \mu \end{pmatrix} : \mu \in \mathbb{R} \right\}$

5. First

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = -2\,.$$

All the other determinants vanish because the first three rows of each matrix are not independent. Indeed, $2R_2 - R_1 = R_3$ in each case, so we can make row operations to get a row of zeros and thus a zero determinant.

6. If U spans $\mathbb{R}^3$, then we must be able to express any vector $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ as

$$X = c^1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + c^2 \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} + c^3 \begin{pmatrix} a \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & a \\ 0 & 2 & 1 \\ 1 & -3 & 0 \end{pmatrix} \begin{pmatrix} c^1 \\ c^2 \\ c^3 \end{pmatrix},$$

for some coefficients c^1, c^2 and c^3. This is a linear system. We could solve for c^1, c^2 and c^3 using an augmented matrix and row operations. However, since we know that $\dim \mathbb{R}^3 = 3$, if U spans $\mathbb{R}^3$, it will also be a basis. Then the solution for c^1, c^2 and c^3 would be unique. Hence, the 3×3 matrix above must be invertible, so we examine its determinant

$$\det \begin{pmatrix} 1 & 1 & a \\ 0 & 2 & 1 \\ 1 & -3 & 0 \end{pmatrix} = 1.(2.0 - 1.(-3)) + 1.(1.1 - a.2) = 4 - 2a\,.$$

Thus U spans $\mathbb{R}^3$ whenever $a \neq 2$. When $a = 2$ we can write the third vector in U in terms of the preceding ones as

$$\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \frac{3}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}.$$

(*You can obtain this result, or an equivalent one by studying the above linear system with $X = 0$, i.e., the associated homogeneous system.*) The two

vectors $\begin{pmatrix}1\\2\\-3\end{pmatrix}$ and $\begin{pmatrix}2\\1\\0\end{pmatrix}$ are clearly linearly independent, so this is the least number of vectors spanning U for this value of a. Also we see that $\dim U = 2$ in this case. Your picture should be a plane in $\mathbb{R}^3$ though the origin containing the vectors $\begin{pmatrix}1\\2\\-3\end{pmatrix}$ and $\begin{pmatrix}2\\1\\0\end{pmatrix}$.

7.

$$\det\begin{pmatrix}1 & x\\1 & y\end{pmatrix} = y - x\,,$$

$$\det\begin{pmatrix}1 & x & x^2\\1 & y & y^2\\1 & z & z^2\end{pmatrix} = \det\begin{pmatrix}1 & x & x^2\\0 & y-x & y^2-x^2\\0 & z-x & z^2-x^2\end{pmatrix}$$

$$= (y-x)(z^2-x^2) - (y^2-x^2)(z-x) = (y-x)(z-x)(z-y)\,.$$

$$\det\begin{pmatrix}1 & x & x^2 & x^3\\1 & y & y^2 & y^3\\1 & z & z^2 & z^3\\1 & w & w^2 & w^3\end{pmatrix} = \det\begin{pmatrix}1 & x & x^2 & x^3\\0 & y-x & y^2-x^2 & y^3-x^3\\0 & z-x & z^2-x^2 & z^3-x^3\\0 & w-x & w^2-x^2 & w^3-x^3\end{pmatrix}$$

$$= \det\begin{pmatrix}1 & 0 & 0 & 0\\0 & y-x & y(y-x) & y^2(y-x)\\0 & z-x & z(z-x) & z^2(z-x)\\0 & w-x & w(w-x) & w^2(w-x)\end{pmatrix}$$

$$= (y-x)(z-x)(w-x)\det\begin{pmatrix}1 & 0 & 0 & 0\\0 & 1 & y & y^2\\0 & 1 & z & z^2\\0 & 1 & w & w^2\end{pmatrix}$$

$$= (y-x)(z-x)(w-x)\det\begin{pmatrix}1 & y & y^2\\1 & z & z^2\\1 & w & w^2\end{pmatrix}$$

$$= (y-x)(z-x)(w-x)(z-y)(w-y)(w-z)\,.$$

From the 4×4 case above, you can see all the tricks required for a general Vandermonde matrix. First zero out the first column by subtracting the first row from all other rows (which leaves the determinant unchanged).

Now zero out the top row by subtracting x_1 times the first column from the second column, x_1 times the second column from the third column *et cetra*. Again these column operations do not change the determinant. Now factor out $x_2 - x_1$ from the second row, $x_3 - x_1$ from the third row, *etc*. This does change the determinant so we write these factors outside the remaining determinant, which is just the same problem but for the $(n-1)\times(n-1)$ case. Iterating the same procedure gives the result

$$\det\begin{pmatrix} 1 & x_1 & (x_1)^2 & \cdots & (x_1)^{n-1} \\ 1 & x_2 & (x_2)^2 & \cdots & (x_2)^{n-1} \\ 1 & x_3 & (x_3)^2 & \cdots & (x_3)^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & (x_n)^2 & \cdots & (x_n)^{n-1} \end{pmatrix} = \prod_{i>j}(x_i - x_j)\,.$$

(Here $\prod$ stands for a multiple product, just like Σ stands for a multiple sum.)

8. (a) No, a basis for $\mathbb{R}^3$ must have exactly three vectors.

   (b) We first extend the original vectors by the standard basis for $\mathbb{R}^4$ and then try to eliminate two of them by considering

$$\alpha\begin{pmatrix}1\\2\\3\\4\end{pmatrix} + \beta\begin{pmatrix}4\\3\\2\\1\end{pmatrix} + \gamma\begin{pmatrix}1\\0\\0\\0\end{pmatrix} + \delta\begin{pmatrix}0\\1\\0\\0\end{pmatrix} + \varepsilon\begin{pmatrix}0\\0\\1\\0\end{pmatrix} + \eta\begin{pmatrix}0\\0\\0\\1\end{pmatrix} = 0\,.$$

   So we study

$$\begin{pmatrix} 1 & 4 & 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 1 & 0 & 0 \\ 3 & 2 & 0 & 0 & 1 & 0 \\ 4 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 4 & 1 & 0 & 0 & 0 \\ 0 & -5 & -2 & 1 & 0 & 0 \\ 0 & -10 & -3 & 0 & 1 & 0 \\ 0 & -15 & -4 & 0 & 0 & 1 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 0 & -\frac{3}{5} & -4 & 0 & 0 \\ 0 & 1 & \frac{2}{5} & \frac{1}{5} & 0 & 0 \\ 0 & 0 & 1 & 10 & 1 & 0 \\ 0 & 0 & 2 & 15 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 2 & \frac{3}{5} & 0 \\ 0 & 1 & 0 & -\frac{19}{5} & -\frac{2}{5} & 0 \\ 0 & 0 & 1 & 10 & 1 & 0 \\ 0 & 0 & 0 & -\frac{5}{2} & -10 & \frac{1}{2} \end{pmatrix}$$

   From here we can keep row reducing to achieve RREF, but we can already see that the non-pivot variables will be ε and η. Hence we can

eject the last two vectors and obtain as our basis

$$\left\{ \begin{pmatrix}1\\2\\3\\4\end{pmatrix}, \begin{pmatrix}4\\3\\2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\0\end{pmatrix} \right\}.$$

Of course, this answer is far from unique!

(c) The method is the same as above. Add the standard basis to $\{u, v\}$ to obtain the linearly dependent set $\{u, v, e_1, \ldots, e_n\}$. Then put these vectors as the columns of a matrix and row reduce. The standard basis vectors in columns corresponding to the non-pivot variables can be removed.

9. (a)

$$\det \begin{pmatrix} \lambda & -\frac{1}{2} & -1 \\ -\frac{1}{2} & \lambda-\frac{1}{2} & -\frac{1}{2} \\ -1 & -\frac{1}{2} & \lambda \end{pmatrix} = \lambda\Big((\lambda-\frac{1}{2})\lambda-\frac{1}{4}\Big)+\frac{1}{2}\Big(-\frac{\lambda}{2}-\frac{1}{2}\Big)-\Big(-\frac{1}{4}+\lambda\Big)$$

$$= \lambda^3 - \frac{1}{2}\lambda^2 - \frac{3}{2}\lambda = \lambda(\lambda+1)(\lambda-\frac{3}{2}).$$

Hence the eigenvalues are $0, -1, \frac{3}{2}$.

(b) When $\lambda = 0$ we must solve the homogenous system

$$\left(\begin{array}{ccc|c} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 0 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

So we find the eigenvector $\begin{pmatrix} s \\ -2s \\ s \end{pmatrix}$ where $s \neq 0$ is arbitrary.

For $\lambda = -1$

$$\left(\begin{array}{ccc|c} 1 & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

So we find the eigenvector $\begin{pmatrix} -s \\ 0 \\ s \end{pmatrix}$ where $s \neq 0$ is arbitrary.

Finally, for $\lambda = \frac{3}{2}$

$$\left(\begin{array}{ccc|c} -\frac{3}{2} & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & -\frac{3}{2} & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & \frac{1}{2} & -\frac{3}{2} & 0 \\ 0 & -\frac{5}{4} & \frac{5}{4} & 0 \\ 0 & \frac{5}{4} & -\frac{5}{4} & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

So we find the eigenvector $\begin{pmatrix} s \\ s \\ s \end{pmatrix}$ where $s \neq 0$ is arbitrary.

If the mistake X is in the direction of the eigenvector $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$, then $Y = 0$. *I.e.*, the satellite returns to the origin $\mathcal{O}$. For all subsequent orbits it will again return to the origin. NASA would be very pleased in this case.

If the mistake X is in the direction $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$, then $Y = -X$. Hence the satellite will move to the point opposite to X. After next orbit will move back to X. It will continue this wobbling motion indefinitely. Since this is a stable situation, again, the elite engineers will pat themselves on the back.

Finally, if the mistake X is in the direction $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, the satellite will move to a point $Y = \frac{3}{2}X$ which is further away from the origin. The same will happen for all subsequent orbits, with the satellite moving a factor $3/2$ further away from $\mathcal{O}$ each orbit (in reality, after several orbits, the approximations used by the engineers in their calculations probably fail and a new computation will be needed). In this case, the satellite will be lost in outer space and the engineers will likely lose their jobs!

10. (a) A basis for B^3 is $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$

    (b) 3.

    (c) $2^3 = 8$.

    (d) $\dim B^3 = 3$.

    (e) Because the vectors $\{v_1, v_2, v_3\}$ are a basis any element $v \in B^3$ can be written uniquely as $v = b^1 v_1 + b^2 v_2 + b^3 v_3$ for some triplet of bits $\begin{pmatrix} b^1 \\ b^2 \\ b^3 \end{pmatrix}$.

Hence, to compute $L(v)$ we use linearity of L

$$L(v) = L(b^1v_1 + b^2v_2 + b^3v_3) = b^1L(v_1) + b^2L(v_2) + b^3L(v_3)$$

$$= \begin{pmatrix} L(v_1) & L(v_2) & L(v_3) \end{pmatrix} \begin{pmatrix} b^1 \\ b^2 \\ b^3 \end{pmatrix} .$$

(f) From the notation of the previous part, we see that we can list linear transformations $L : B^3 \to B$ by writing out all possible bit-valued row vectors

$$\begin{pmatrix} 0 & 0 & 0 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix},$$
$$\begin{pmatrix} 0 & 1 & 0 \end{pmatrix},$$
$$\begin{pmatrix} 0 & 0 & 1 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 1 & 0 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 & 1 \end{pmatrix},$$
$$\begin{pmatrix} 0 & 1 & 1 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

There are $2^3 = 8$ different linear transformations $L : B^3 \to B$, exactly the same as the number of elements in B^3.

(g) Yes, essentially just because L_1 and L_2 are linear transformations. In detail for any bits (a, b) and vectors (u, v) in B^3 it is easy to check the linearity property for $(\alpha L_1 + \beta L_2)$

$$(\alpha L_1 + \beta L_2)(au + bv) = \alpha L_1(au + bv) + \beta L_2(au + bv)$$

$$= \alpha a L_1(u) + \alpha b L_1(v) + \beta a L_1(u) + \beta b L_1(v)$$

$$= a(\alpha L_1(u) + \beta L_2(v)) + b(\alpha L_1(u) + \beta L_2(v))$$

$$= a(\alpha L_1 + \beta L_2)(u) + b(\alpha L_1 + \beta L_2)(v) .$$

Here the first line used the definition of $(\alpha L_1 + \beta L_2)$, the second line depended on the linearity of L_1 and L_2, the third line was just algebra and the fourth used the definition of $(\alpha L_1 + \beta L_2)$ again.

(h) Yes. The easiest way to see this is the identification above of these maps with bit-valued column vectors. In that notation, a basis is

$$\Big\{ \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \Big\} .$$

Since this (spanning) set has three (linearly independent) elements, the vector space of linear maps $B^3 \to B$ has dimension 3. This is an example of a general notion called the *dual vector space*.

11. (a) $\frac{d^2 X}{dt^2} = \frac{d^2 \cos(\omega t)}{dt^2} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = -\omega^2 \cos(\omega t) \begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

Hence

$$
\begin{aligned}
F = \cos(\omega t) \begin{pmatrix} -a - b \\ -a - 2b - c \\ -b - c \end{pmatrix} &= \cos(\omega t) \begin{pmatrix} -1 & -1 & 0 \\ -1 & -2 & -1 \\ 0 & -1 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \\
&= -\omega^2 \cos(\omega t) \begin{pmatrix} a \\ b \\ c \end{pmatrix} ,
\end{aligned}
$$

so

$$
M = \begin{pmatrix} -1 & -1 & 0 \\ -1 & -2 & -1 \\ 0 & -1 & -1 \end{pmatrix} .
$$

(b)

$$
\begin{aligned}
\det \begin{pmatrix} \lambda + 1 & 1 & 0 \\ 1 & \lambda + 2 & 1 \\ 0 & 1 & \lambda + 1 \end{pmatrix} &= (\lambda + 1)\big((\lambda + 2)(\lambda + 1) - 1\big) - (\lambda + 1) \\
&= (\lambda + 1)\big((\lambda + 2)(\lambda + 1) - 2\big) \\
&= (\lambda + 1)\big(\lambda^2 + 3\lambda\big) = \lambda(\lambda + 1)(\lambda + 3)
\end{aligned}
$$

so the eigenvalues are $\lambda = 0, -1, -3$.

For the eigenvectors, when $\lambda = 0$ we study:

$$
M - 0.I = \begin{pmatrix} -1 & -1 & 0 \\ -1 & -2 & -1 \\ 0 & -1 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & -1 \\ 0 & -1 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} ,
$$

so $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ is an eigenvector.

For $\lambda = -1$

$$
M - (-1).I = \begin{pmatrix} 0 & -1 & 0 \\ -1 & -1 & -1 \\ 0 & -1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} ,
$$

so $\begin{pmatrix}-1\\0\\1\end{pmatrix}$ is an eigenvector.

For $\lambda=-3$

$$M-(-3).I=\begin{pmatrix}2&-1&0\\-1&1&-1\\0&-1&2\end{pmatrix}\sim\begin{pmatrix}1&-1&1\\0&1&-2\\0&-1&2\end{pmatrix}\sim\begin{pmatrix}1&0&-1\\0&1&-2\\0&0&0\end{pmatrix},$$

so $\begin{pmatrix}1\\2\\1\end{pmatrix}$ is an eigenvector.

(c) The characteristic frequencies are $0, 1, \sqrt{3}$.

(d) The orthogonal change of basis matrix

$$P=\begin{pmatrix}\frac{1}{\sqrt{3}}&-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{6}}\\-\frac{1}{\sqrt{3}}&0&\frac{2}{\sqrt{6}}\\\frac{1}{\sqrt{3}}&\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{6}}\end{pmatrix}$$

It obeys $MP=PD$ where

$$D=\begin{pmatrix}0&0&0\\0&-1&0\\0&0&-3\end{pmatrix}.$$

(e) Yes, the direction given by the eigenvector $\begin{pmatrix}1\\-1\\1\end{pmatrix}$ because its eigenvalue is zero. This is probably a bad design for a bridge because it can be displaced in this direction with no force!

12. (a) If we call $M=\begin{pmatrix}a&b\\b&d\end{pmatrix}$, then $X^TMX=ax^2+2bxy+dy^2$. Similarly putting $C=\begin{pmatrix}c\\e\end{pmatrix}$ yields $X^TC+C^TX=2X\cdot C=2cx+2ey$. Thus

$$0=ax^2+2bxy+dy^2+2cx+2ey+f$$

$$=\begin{pmatrix}x&y\end{pmatrix}\begin{pmatrix}a&b\\b&d\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+\begin{pmatrix}x&y\end{pmatrix}\begin{pmatrix}c\\e\end{pmatrix}+\begin{pmatrix}c&e\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}+f.$$

(b) Yes, the matrix M is symmetric, so it will have a basis of eigenvectors and is similar to a diagonal matrix of real eigenvalues.

To find the eigenvalues notice that $\det\begin{pmatrix} a-\lambda & b \\ b & d-\lambda\end{pmatrix} = (a-\lambda)(d-\lambda) - b^2 = \big(\lambda - \frac{a+d}{2}\big)^2 - b^2 - \big(\frac{a-d}{2}\big)^2$. So the eigenvalues are

$$\lambda = \frac{a+d}{2} + \sqrt{b^2 + \big(\frac{a-d}{2}\big)^2} \text{ and } \mu = \frac{a+d}{2} - \sqrt{b^2 + \big(\frac{a-d}{2}\big)^2}\,.$$

(c) The trick is to write

$$X^T M X + C^T X + X^T C = (X^T + C^T M^{-1}) M (X + M^{-1} C) - C^T M^{-1} C\,,$$

so that

$$(X^T + C^T M^{-1}) M (X + M^{-1} C) = C^T M C - f\,.$$

Hence $Y = X + M^{-1}C$ and $g = C^T M C - f$.

(d) The cosine of the angle between vectors V and W is given by

$$\frac{V \cdot W}{\sqrt{V \cdot V\, W \cdot W}} = \frac{V^T W}{\sqrt{V^T V\, W^T W}}\,.$$

So replacing $V \to PV$ and $W \to PW$ will always give a factor $P^T P$ inside all the products, but $P^T P = I$ for orthogonal matrices. Hence none of the dot products in the above formula changes, so neither does the angle between V and W.

(e) If we take the eigenvectors of M, normalize them (*i.e.* divide them by their lengths), and put them in a matrix P (as columns) then P will be an orthogonal matrix. (If it happens that $\lambda = \mu$, then we also need to make sure the eigenvectors spanning the two dimensional eigenspace corresponding to λ are orthogonal.) Then, since M times the eigenvectors yields just the eigenvectors back again multiplied by their eigenvalues, it follows that $MP = PD$ where D is the diagonal matrix made from eigenvalues.

(f) If $Y = PZ$, then $Y^T M Y = Z^T P^T M P Z = Z^T P^T P D Z = Z^T D Z$ where $D = \begin{pmatrix}\lambda & 0 \\ 0 & \mu\end{pmatrix}$.

(g) Using part (f) and (c) we have

$$\lambda z^2 + \mu w^2 = g\,.$$

(h) When $\lambda = \mu$ and $g/\lambda = R^2$, we get the equation for a circle radius R in the (z, w)-plane. When λ, μ and g are postive, we have the equation for an ellipse. Vanishing g along with λ and μ of opposite signs gives a pair of straight lines. When g is non-vanishing, but λ and μ have opposite signs, the result is a pair of hyperbolæ. These shapes all come from cutting a cone with a plane, and are therefore called conic sections.

13. We show that L is bijective if and only if M is invertible.

(a) We suppose that L is bijective.

i. Since L is injective, its kernel consists of the zero vector alone. Hence
$$L = \dim \ker L = 0.$$
So by the Dimension Formula,
$$\dim V = L + \operatorname{rank} L = \operatorname{rank} L.$$
Since L is surjective, $L(V) = W$. Thus
$$\operatorname{rank} L = \dim L(V) = \dim W.$$
Thereby
$$\dim V = \operatorname{rank} L = \dim W.$$

ii. Since $\dim V = \dim W$, the matrix M is square so we can talk about its eigenvalues. Since L is injective, its kernel is the zero vector alone. That is, the only solution to $LX = 0$ is $X = 0_V$. But LX is the same as MX, so the only solution to $MX = 0$ is $X = 0_V$. So M does not have zero as an eigenvalue.

iii. Since $MX = 0$ has no non-zero solutions, the matrix M is invertible.

(b) Now we suppose that M is an invertible matrix.

i. Since M is invertible, the system $MX = 0$ has no non-zero solutions. But LX is the same as MX, so the only solution to $LX = 0$ is $X = 0_V$. So L does not have zero as an eigenvalue.

ii. Since $LX = 0$ has no non-zero solutions, the kernel of L is the zero vector alone. So L is injective.

iii. Since M is invertible, we must have that $\dim V = \dim W$. By the Dimension Formula, we have
$$\dim V = L + \operatorname{rank} L$$

and since $\ker L = \{0_V\}$ we have $L = \dim \ker L = 0$, so

$$\dim W = \dim V = \operatorname{rank} L = \dim L(V).$$

Since $L(V)$ is a subspace of W with the same dimension as W, it must be equal to W. To see why, pick a basis B of $L(V)$. Each element of B is a vector in W, so the elements of B form a linearly independent set in W. Therefore B is a basis of W, since the size of B is equal to $\dim W$. So $L(V) = \operatorname{span} B = W$. So L is surjective.

14. (a) $F_4 = F_2 + F_3 = 2 + 3 = 5$.

(b) The number of pairs of doves in any given year equals the number of the previous years plus those that hatch and there are as many of them as pairs of doves in the year before the previous year.

(c) $X_1 = \begin{pmatrix} F_1 \\ F_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $X_2 = \begin{pmatrix} F_2 \\ F_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

$$MX_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = X_2\,.$$

(d) We just need to use the recursion relationship of part (b) in the top slot of X_{n+1}:

$$X_{n+1} = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = \begin{pmatrix} F_n + F_{n-1} \\ F_n \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix} = MX_n\,.$$

(e) Notice M is symmetric so this is guaranteed to work.

$$\det \begin{pmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda(\lambda-1) - 1 = \big(\lambda - \frac{1}{2}\big)^2 - \frac{5}{4}\,,$$

so the eigenvalues are $\frac{1\pm\sqrt{5}}{2}$. Hence the eigenvectors are $\begin{pmatrix} \frac{1\pm\sqrt{5}}{2} \\ 1 \end{pmatrix}$, respectively (notice that $\frac{1+\sqrt{5}}{2} + 1 = \frac{1+\sqrt{5}}{2}.\frac{1+\sqrt{5}}{2}$ and $\frac{1-\sqrt{5}}{2} + 1 = \frac{1-\sqrt{5}}{2}.\frac{1-\sqrt{5}}{2}$). Thus $M = PDP^{-1}$ with

$$D = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix} \text{ and } P = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix}.$$

(f) $M^n = (PDP^{-1})^n = PDP^{-1}PDP^{-1}\dots PDP^{-1} = PD^nP^{-1}$.

(g) Just use the matrix recursion relation of part (d) repeatedly:

$$X_{n+1} = MX_n = M^2 X_{n-1} = \cdots = M^n X_1 \,.$$

(h) The eigenvalues are $\varphi = \frac{1+\sqrt{5}}{2}$ and $1-\varphi = \frac{1-\sqrt{5}}{2}$.

(i)

$$X_{n+1} = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = M^n X_n = PD^nP^{-1}X_1$$

$$= P\begin{pmatrix} \varphi & 0 \\ 0 & 1-\varphi \end{pmatrix}^n \begin{pmatrix} \frac{1}{\sqrt{5}} & \star \\ -\frac{1}{\sqrt{5}} & \star \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = P\begin{pmatrix} \varphi^n & 0 \\ 0 & (1-\varphi)^n \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{\varphi^n}{\sqrt{5}} \\ -\frac{(1-\varphi)^n}{\sqrt{5}} \end{pmatrix} = \begin{pmatrix} \star \\ \frac{\varphi^n-(1-\varphi)^n}{\sqrt{5}} \end{pmatrix}.$$

Hence

$$F_n = \frac{\varphi^n - (1-\varphi)^n}{\sqrt{5}} \,.$$

These are the famous Fibonacci numbers.

15. Call the three vectors u, v and w, respectively. Then

$$v^\perp = v - \frac{u \cdot v}{u \cdot u} u = v - \frac{3}{4} u = \begin{pmatrix} \frac{1}{4} \\ -\frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix},$$

and

$$w^\perp = w - \frac{u \cdot w}{u \cdot u} u - \frac{v^\perp \cdot w}{v^\perp \cdot v^\perp} v^\perp = w - \frac{3}{4} u - \frac{\frac{3}{4}}{\frac{3}{4}} v^\perp = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Dividing by lengths, an orthonormal basis for $\operatorname{span}\{u, v, w\}$ is

$$\left\{ \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{\sqrt{3}}{6} \\ -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{6} \\ \frac{\sqrt{3}}{6} \end{pmatrix}, \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ 0 \\ 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix} \right\}.$$

16. (a) The columns of M span $\operatorname{im} L$ in the basis given for W.

(b) The rows of M span $(\ker L)^{\perp}$

(c) First we put M in RREF:

$$M = \begin{pmatrix}1&2&1&3\\2&1&-1&2\\1&0&0&-1\\4&1&-1&0\end{pmatrix} \sim \begin{pmatrix}1&2&1&3\\0&-3&-3&-4\\0&-2&-1&-4\\0&-7&-5&-12\end{pmatrix}$$

$$\sim \begin{pmatrix}1&0&-1&\frac{1}{3}\\0&1&1&\frac{4}{3}\\0&0&1&-\frac{4}{3}\\0&0&2&-\frac{8}{3}\end{pmatrix} \sim \begin{pmatrix}1&0&0&-1\\0&1&0&\frac{8}{3}\\0&0&1&-\frac{4}{3}\\0&0&0&0\end{pmatrix}.$$

Hence

$$\ker L = \operatorname{span}\{v_1 - \frac{8}{3}v_2 + \frac{4}{3}v_3 + v_4\}$$

and

$$\operatorname{im}L = \operatorname{span}\{v_1 + 2v_2 + v_3 + 4v_4, 2v_1 + v_2 + v_4, v_1 - v_2 - v_4\}.$$

Thus $\dim\ker L = 1$ and $\dim \operatorname{im}L = 3$ so

$$\dim\ker L + \dim\operatorname{im}L = 1 + 3 = 4 = \dim V.$$

17. (a)

$$\begin{cases} 5 = 4a - 2b + c \\ 2 = a - b + c \\ 0 = a + b + c \\ 3 = 4a + 2b + c. \end{cases}$$

(b,c,d)

$$\left(\begin{array}{ccc|c}4&-2&1&5\\1&-1&1&2\\1&1&1&0\\4&2&1&3\end{array}\right) \sim \left(\begin{array}{ccc|c}1&1&1&0\\0&-6&-3&5\\0&-2&0&2\\0&-2&-3&3\end{array}\right) \sim \left(\begin{array}{ccc|c}1&0&1&-1\\0&1&0&1\\0&0&-3&11\\0&0&-3&3\end{array}\right)$$

The system has no solutions because $c = -1$ and $c = -\frac{11}{3}$ is impossible.

(e) Let

$$M = \begin{pmatrix}4&-2&1\\1&-1&1\\1&1&1\\4&2&1\end{pmatrix} \text{ and } V = \begin{pmatrix}5\\2\\0\\3\end{pmatrix}.$$

Then

$$M^TM = \begin{pmatrix} 34 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 4 \end{pmatrix} \text{ and } M^TV = \begin{pmatrix} 34 \\ -6 \\ 10 \end{pmatrix}.$$

So

$$\left(\begin{array}{ccc|c} 34 & 0 & 10 & 34 \\ 0 & 10 & 0 & -6 \\ 10 & 0 & 4 & 10 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & \frac{2}{5} & 1 \\ 0 & 10 & 0 & -6 \\ 0 & 0 & -\frac{18}{5} & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -\frac{3}{5} \\ 0 & 0 & 1 & 0 \end{array}\right)$$

The least squares solution is $a = 1$, $b = -\frac{3}{5}$ and $c = 0$.

(b) The Captain predicts $y(2) = 1.2^2 - \frac{3}{5}.2 + 0 = \frac{14}{5}$.

18.

G Movie Scripts

G.1 What is Linear Algebra?

Hint for Review Problem 5

Looking at the problem statement we find some important information, first that oranges always have twice as much sugar as apples, and second that the information about the barrel is recorded as (s, f), where $s =$ units of sugar in the barrel and $f =$ number of pieces of fruit in the barrel.
We are asked to find a linear transformation relating this new representation to the one in the lecture, where in the lecture $x =$ the number of apples and $y =$ the number of oranges. This means we must create a system of equations relating the variable x and y to the variables s and f in matrix form. Your answer should be the matrix that transforms one set of variables into the other.

Hint: Let λ represent the amount of sugar in each apple.

1. To find the first equation relate f to the variables x and y.
2. To find the second equation, use the hint to figure out how much sugar is in x apples, and y oranges in terms of λ. Then write an equation for s using x, y and λ.

G.2 Systems of Linear Equations

Augmented Matrix Notation

Why is the augmented matrix

$$\left(\begin{array}{cc|c} 1 & 1 & 27 \\ 2 & -1 & 0 \end{array}\right),$$

equivalent to the system of equations

$$\begin{aligned} x+y &= 27 \\ 2x-y &= 0\,? \end{aligned}$$

Well the augmented matrix is just a new notation for the matrix equation

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 27 \\ 0 \end{pmatrix}$$

and if you review your matrix multiplication remember that

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ 2x-y \end{pmatrix}$$

This means that

$$\begin{pmatrix} x+y \\ 2x-y \end{pmatrix} = \begin{pmatrix} 27 \\ 0 \end{pmatrix},$$

which is our original equation.

Equivalence of Augmented Matrices

Lets think about what it means for the two augmented matrices

$$\left(\begin{array}{cc|c} 1 & 1 & 27 \\ 2 & -1 & 0 \end{array}\right) \text{ and } \left(\begin{array}{cc|c} 1 & 0 & 9 \\ 0 & 1 & 18 \end{array}\right)$$

to be equivalent: They are certainly not equal, because they don't match in each component, but since these augmented matrices represent a system, we might want to introduce a new kind of equivalence relation.

Well we could look at the system of linear equations this represents

$$\begin{aligned} x+y &= 27 \\ 2x-y &= 0 \end{aligned}$$

and notice that the solution is $x=9$ and $y=18$. The other augmented matrix represents the system

$$\begin{aligned} x \;+ 0\cdot y &= 9 \\ 0\cdot x \;+ \quad y &= 18 \end{aligned}$$

This clearly has the same solution. The first and second system are related in the sense that their solutions are the same. Notice that it is really nice to have the augmented matrix in the second form, because the matrix multiplication can be done in your head.

Hints for Review Question 10

This question looks harder than it actually is:

> Row equivalence of matrices is an example of an *equivalence relation*. Recall that a relation $\sim$ on a set of objects U is an equivalence relation if the following three properties are satisfied:
>
> - Reflexive: For any $x \in U$, we have $x \sim x$.
> - Symmetric: For any $x, y \in U$, if $x \sim y$ then $y \sim x$.
> - Transitive: For any x, y and $z \in U$, if $x \sim y$ and $y \sim z$ then $x \sim z$.
>
> (For a more complete discussion of equivalence relations, see Webwork Homework 0, Problem 4)
>
> Show that row equivalence of augmented matrices is an equivalence relation.

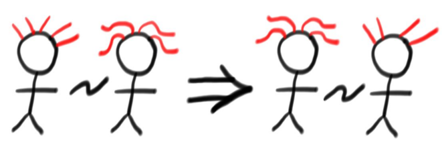
Firstly remember that an equivalence relation is just a more general version of ''equals''. Here we defined row equivalence for augmented matrices whose linear systems have solutions by the property that their solutions are the same.

So this question is really about the word *same*. Lets do a silly example: Lets replace the set of augmented matrices by the set of people who have hair. We will call two people equivalent if they have the same hair color. There are three properties to check:

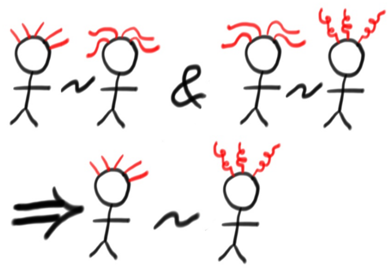
- Reflexive: This just requires that you have the same hair color as yourself so obviously holds.



- Symmetric: If the first person, Bob (say) has the same hair color as a second person Betty(say), then Bob has the same hair color as Betty, so this holds too.



- Transitive: If Bob has the same hair color as Betty (say) and Betty has the same color as Brenda (say), then it follows that Bob and Brenda have the same hair color, so the transitive property holds too and we are done.



Solution set in set notation

Here is an augmented matrix, let's think about what the solution set looks like

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 2 \\ 0 & 1 & 0 & 1 \end{array}\right)$$

This looks like the system

$$\begin{array}{rcl} 1 \cdot x_1 \quad + \; 3x_3 & = & 2 \\ 1 \cdot x_2 \qquad & = & 1 \end{array}$$

Notice that when the system is written this way the copy of the 2×2 identity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ makes it easy to write a solution in terms of the variables x_1 and x_2. We will call x_1 and x_2 the *pivot* variables. The third column $\begin{pmatrix} 3 \\ 0 \end{pmatrix}$ does not look like part of an identity matrix, and there is no 3×3 identity in the augmented matrix. Notice there are more variables than equations and that this means we will have to write the solutions for the system in terms of the variable x_3. We'll call x_3 the *free* variable.

Let $x_3 = \mu$. (We could also just add a ''dummy'' equation $x_3 = x_3$.) Then we can rewrite the first equation in our system

$$\begin{array}{rcl} x_1 + 3x_3 & = & 2 \\ x_1 + 3\mu & = & 2 \\ x_1 & = & 2 - 3\mu. \end{array}$$

Then since the second equation doesn't depend on μ we can keep the equation

$$x_2 = 1,$$

and for a third equation we can write

$$x_3 = \mu$$

so that we get the system

$$\begin{array}{rcl} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} & = & \begin{pmatrix} 2 - 3\mu \\ 1 \\ \mu \end{pmatrix} \\ & = & \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -3\mu \\ 0 \\ \mu \end{pmatrix} \\ & = & \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix}. \end{array}$$

Any value of μ will give a solution of the system, and any system can be written in this form for some value of μ. Since there are multiple solutions, we can also express them as a set:

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \middle| \mu \in \mathbb{R} \right\}.$$

Worked Examples of Gaussian Elimination

Let us consider that we are given two systems of equations that give rise to the following two (augmented) matrices:

$$\left(\begin{array}{cccc|c} 2 & 5 & 2 & 0 & 2 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 4 & 1 & 0 & 1 \end{array}\right) \qquad \left(\begin{array}{cc|c} 5 & 2 & 9 \\ 0 & 5 & 10 \\ 0 & 3 & 6 \end{array}\right)$$

and we want to find the solution to those systems. We will do so by doing Gaussian elimination.

For the first matrix we have

$$\left(\begin{array}{cccc|c} 2 & 5 & 2 & 0 & 2 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 4 & 1 & 0 & 1 \end{array}\right) \overset{R_1 \leftrightarrow R_2}{\sim} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 0 & 1 \\ 2 & 5 & 2 & 0 & 2 \\ 1 & 4 & 1 & 0 & 1 \end{array}\right)$$

$$\overset{R_2 - 2R_1; R_3 - R_1}{\sim} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 0 & 1 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \end{array}\right)$$

$$\overset{\frac{1}{3}R_2}{\sim} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \end{array}\right)$$

$$\overset{R_1 - R_2; R_3 - 3R_2}{\sim} \left(\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

1. We begin by interchanging the first two rows in order to get a 1 in the upper-left hand corner and avoiding dealing with fractions.
2. Next we subtract row 1 from row 3 and twice from row 2 to get zeros in the left-most column.
3. Then we scale row 2 to have a 1 in the eventual pivot.
4. Finally we subtract row 2 from row 1 and three times from row 2 to get it into Reduced Row Echelon Form.

Therefore we can write $x = 1 - \lambda$, $y = 0$, $z = \lambda$ and $w = \mu$, or in vector form

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Now for the second system we have

$$\left(\begin{array}{cc|c} 5 & 2 & 9 \\ 0 & 5 & 10 \\ 0 & 3 & 6 \end{array}\right) \stackrel{\frac{1}{5}R_2}{\sim} \left(\begin{array}{cc|c} 5 & 2 & 9 \\ 0 & 1 & 2 \\ 0 & 3 & 6 \end{array}\right)$$
$$\stackrel{R_3-3R_2}{\sim} \left(\begin{array}{cc|c} 5 & 2 & 9 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right)$$
$$\stackrel{R_1-2R_2}{\sim} \left(\begin{array}{cc|c} 5 & 0 & 5 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right)$$
$$\stackrel{\frac{1}{5}R_1}{\sim} \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right)$$

We scale the second and third rows appropriately in order to avoid fractions, then subtract the corresponding rows as before. Finally scale the first row and hence we have $x = 1$ and $y = 2$ as a unique solution.

Hints for Review Question 10

This question looks harder than it actually is:

> Row equivalence of matrices is an example of an *equivalence relation*. Recall that a relation $\sim$ on a set of objects U is an equivalence relation if the following three properties are satisfied:
>
> - Reflexive: For any $x \in U$, we have $x \sim x$.
> - Symmetric: For any $x, y \in U$, if $x \sim y$ then $y \sim x$.
> - Transitive: For any x, y and $z \in U$, if $x \sim y$ and $y \sim z$ then $x \sim z$.
>
> (For a more complete discussion of equivalence relations, see Webwork Homework 0, Problem 4)
>
> Show that row equivalence of augmented matrices is an equivalence relation.

Firstly remember that an equivalence relation is just a more general version of ‘‘equals’’. Here we defined row equivalence for augmented matrices whose linear systems have solutions by the property that their solutions are the same.

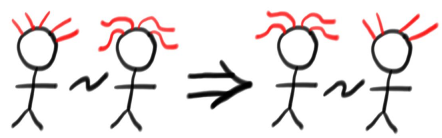
So this question is really about the word *same*. Lets do a silly example: Lets replace the set of augmented matrices by the set of people who have hair.

We will call two people equivalent if they have the same hair color. There are three properties to check:

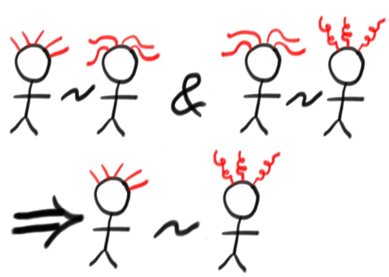
- Reflexive: This just requires that you have the same hair color as yourself so obviously holds.



- Symmetric: If the first person, Bob (say) has the same hair color as a second person Betty(say), then Bob has the same hair color as Betty, so this holds too.



- Transitive: If Bob has the same hair color as Betty (say) and Betty has the same color as Brenda (say), then it follows that Bob and Brenda have the same hair color, so the transitive property holds too and we are done.



Hint for Review Question 5

The first part for Review Question 5 is simple--just write out the associated linear system and you will find the equation $0 = 6$ which is inconsistent. Therefore we learn that we must avoid a row of zeros preceding a non-vanishing entry after the vertical bar.

Turning to the system of equations, we first write out the augmented matrix and then perform two row operations

$$\left(\begin{array}{ccc|c} 1 & -3 & 0 & 6 \\ 1 & 0 & 3 & -3 \\ 2 & k & 3-k & 1 \end{array}\right)$$

$$\stackrel{R_2-R_1;R_3-2R_1}{\sim} \left(\begin{array}{ccc|c} 1 & -3 & 0 & 6 \\ 0 & 3 & 3 & -9 \\ 0 & k+6 & 3-k & -11 \end{array}\right).$$

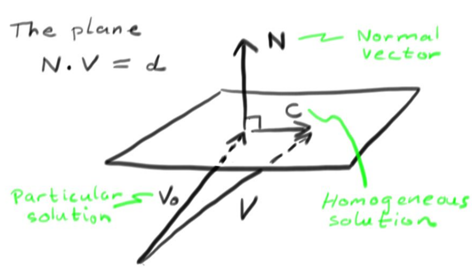
Next we would like to subtract some amount of R_2 from R_3 to achieve a zero in the third entry of the second column. But if

$$k + 6 = 3 - k \Rightarrow k = -\frac{3}{2},$$

this would produce zeros in the third row before the vertical line. You should also check that this does not make the whole third line zero. You now have enough information to write a complete solution.

Planes

Here we want to describe the mathematics of planes in space. The video is summarised by the following picture:



A plane is often called $\mathbb{R}^2$ because it is spanned by two coordinates, and space is called $\mathbb{R}^3$ and has three coordinates, usually called (x, y, z). The equation for a plane is

$$ax + by + cz = d.$$

Lets simplify this by calling $V = (x, y, z)$ the vector of unknowns and $N = (a, b, c)$. Using the dot product in $\mathbb{R}^3$ we have

$$N \cdot V = d\,.$$

Remember that when vectors are perpendicular their dot products vanish. *I.e.* $U \cdot V = 0 \Leftrightarrow U \perp V$. This means that if a vector V_0 solves our equation $N \cdot V = d$, then so too does $V_0 + C$ whenever C is perpendicular to N. This is because

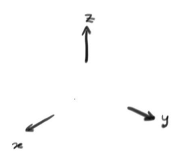
$$N \cdot (V_0 + C) = N \cdot V_0 + N \cdot C = d + 0 = d\,.$$

But C is ANY vector perpendicular to N, so all the possibilities for C span a plane whose normal vector is N. Hence we have shown that solutions to the equation $ax + by + cz = 0$ are a plane with normal vector $N = (a, b, c)$.

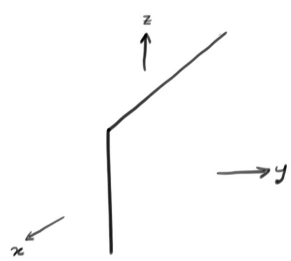
Pictures and Explanation

This video considers solutions sets for linear systems with three unknowns. These are often called (x, y, z) and label points in $\mathbb{R}^3$. Lets work case by case:

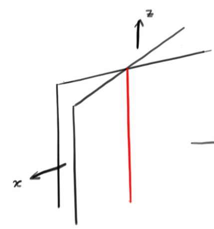
- If you have no equations at all, then any (x, y, z) is a solution, so the solution set is all of $\mathbb{R}^3$. The picture looks a little silly:



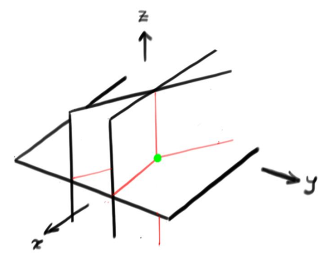
- For a single equation, the solution is a plane. This is explained in this video or the accompanying script. The picture looks like this:



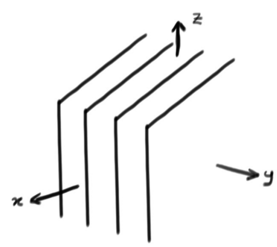
- For two equations, we must look at two planes. These usually intersect along a line, so the solution set will also (usually) be a line:



- For three equations, most often their intersection will be a single point so the solution will then be unique:



- Of course stuff can go wrong. Two different looking equations could determine the same plane, or worse equations could be inconsistent. If the equations are inconsistent, there will be no solutions at all. For example, if you had four equations determining four parallel planes the solution set would be empty. This looks like this:



G.3 Vectors in Space n-Vectors

Review of Parametric Notation

The equation for a plane in three variables x, y and z looks like

$$ax + by + cz = d$$

where a, b, c, and d are constants. Lets look at the example

$$x + 2y + 5z = 3 .$$

In fact this is a system of linear equations whose solutions form a plane with normal vector $(1, 2, 5)$. As an augmented matrix the system is simply

$$\left(\begin{array}{ccc|c} 1 & 2 & 5 & 3 \end{array}\right).$$

This is actually RREF! So we can let x be our pivot variable and y, z be represented by free parameters λ_1 and λ_2:

$$x = \lambda_1 , \qquad y = \lambda_2 .$$

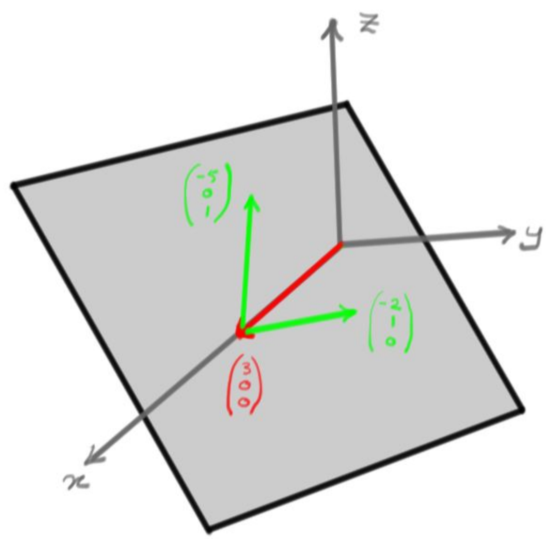
Thus we write the solution as

$$\begin{array}{rcrrr} x & = & -2\lambda_1 & -5\lambda_2 & +3 \\ y & = & \lambda_1 & & \\ z & = & & \lambda_2 & \end{array}$$

or in vector notation

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix} .$$

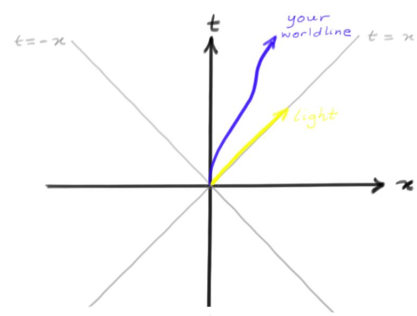
This describes a plane parametric equation. Planes are ‘‘two-dimensional’’ because they are described by two free variables. Here’s a picture of the resulting plane:



The Story of Your Life

This video talks about the weird notion of a ‘‘length-squared’’ for a vector $v = (x, t)$ given by $||v||^2 = x^2 - t^2$ used in Einstein’s theory of relativity. The

idea is to plot the story of your life on a plane with coordinates (x,t). The coordinate x encodes *where* an event happened (for real life situations, we must replace $x \to (x,y,z) \in \mathbb{R}^3$). The coordinate t says *when* events happened. Therefore you can plot your life history as a worldline as shown:



Each point on the worldline corresponds to a place and time of an event in your life. The slope of the worldline has to do with your speed. Or to be precise, the inverse slope is your velocity. Einstein realized that the maximum speed possible was that of light, often called c. In the diagram above $c = 1$ and corresponds to the lines $x = \pm t \Rightarrow x^2 - t^2 = 0$. This should get you started in your search for vectors with zero length.

G.4 Vector Spaces

Examples of Each Rule

Lets show that $\mathbb{R}^2$ is a vector space. To do this (unless we invent some clever tricks) we will have to check all parts of the definition. Its worth doing this once, so here we go:

Before we start, remember that for $\mathbb{R}^2$ we define vector addition and scalar multiplication component-wise.

(+i) Additive closure: We need to make sure that when we add $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ that we do not get something outside the original vector space $\mathbb{R}^2$. This just relies on the underlying structure of real numbers whose sums are again real numbers so, using our component-wise addition law we have

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} := \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} \in \mathbb{R}^2 .$$

(+ii) Additive commutativity: We want to check that when we add any two vectors we can do so in either order, *i.e.*

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \stackrel{?}{=} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} .$$

This again relies on the underlying real numbers which for any $x, y \in \mathbb{R}$ obey

$$x + y = y + x\,.$$

This fact underlies the middle step of the following computation

$$\begin{pmatrix}x_1\\x_2\end{pmatrix} + \begin{pmatrix}y_1\\y_2\end{pmatrix} = \begin{pmatrix}x_1+y_1\\x_2+y_2\end{pmatrix} = \begin{pmatrix}y_1+x_1\\y_2+x_2\end{pmatrix} = \begin{pmatrix}y_1\\y_2\end{pmatrix} + \begin{pmatrix}x_1\\x_2\end{pmatrix}\,,$$

which demonstrates what we wished to show.

(+iii) Additive Associativity: This shows that we needn't specify with parentheses which order we intend to add triples of vectors because their sums will agree for either choice. What we have to check is

$$\left(\begin{pmatrix}x_1\\x_2\end{pmatrix} + \begin{pmatrix}y_1\\y_2\end{pmatrix}\right) + \begin{pmatrix}z_1\\z_2\end{pmatrix} \stackrel{?}{=} \begin{pmatrix}x_1\\x_2\end{pmatrix} + \left(\begin{pmatrix}y_1\\y_2\end{pmatrix} + \begin{pmatrix}z_1\\z_2\end{pmatrix}\right)\,.$$

Again this relies on the underlying associativity of real numbers:

$$(x + y) + z = x + (y + z)\,.$$

The computation required is

$$\left(\begin{pmatrix}x_1\\x_2\end{pmatrix} + \begin{pmatrix}y_1\\y_2\end{pmatrix}\right) + \begin{pmatrix}z_1\\z_2\end{pmatrix} = \begin{pmatrix}x_1+y_1\\x_2+y_2\end{pmatrix} + \begin{pmatrix}z_1\\z_2\end{pmatrix} = \begin{pmatrix}(x_1+y_1)+z_1\\(x_2+y_2)+z_2\end{pmatrix}$$

$$= \begin{pmatrix}x_1+(y_1+z_1)\\x_2+(y_2+z_2)\end{pmatrix} = \begin{pmatrix}x_1\\y_1\end{pmatrix} + \begin{pmatrix}y_1+z_1\\y_2+z_2\end{pmatrix} = \begin{pmatrix}x_1\\x_2\end{pmatrix} + \left(\begin{pmatrix}y_1\\y_2\end{pmatrix} + \begin{pmatrix}z_1\\z_2\end{pmatrix}\right)\,.$$

(iv) Zero: There needs to exist a vector $\vec{0}$ that works the way we would expect zero to behave, *i.e.*

$$\begin{pmatrix}x_1\\y_1\end{pmatrix} + \vec{0} = \begin{pmatrix}x_1\\y_1\end{pmatrix}\,.$$

It is easy to find, the answer is

$$\vec{0} = \begin{pmatrix}0\\0\end{pmatrix}\,.$$

You can easily check that when this vector is added to any vector, the result is unchanged.

(+v) Additive Inverse: We need to check that when we have $\begin{pmatrix}x_1\\x_2\end{pmatrix}$, there is another vector that can be added to it so the sum is $\vec{0}$. (Note that it is important to first figure out what $\vec{0}$ is here!) The answer for the additive inverse of $\begin{pmatrix}x_1\\x_2\end{pmatrix}$ is $\begin{pmatrix}-x_1\\-x_2\end{pmatrix}$ because

$$\begin{pmatrix}x_1\\x_2\end{pmatrix} + \begin{pmatrix}-x_1\\-x_2\end{pmatrix} = \begin{pmatrix}x_1-x_1\\x_2-x_2\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} = \vec{0}\,.$$

We are half-way done, now we need to consider the rules for scalar multiplication. Notice, that we multiply vectors by scalars (*i.e.* numbers) but do NOT multiply a vectors by vectors.

(·i) Multiplicative closure: Again, we are checking that an operation does not produce vectors outside the vector space. For a scalar $a \in \mathbb{R}$, we require that $a \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ lies in $\mathbb{R}^2$. First we compute using our component-wise rule for scalars times vectors:

$$a \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \end{pmatrix} .$$

Since products of real numbers ax_1 and ax_2 are again real numbers we see this is indeed inside $\mathbb{R}^2$.

(·ii) Multiplicative distributivity: The equation we need to check is

$$(a+b) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \stackrel{?}{=} a \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} .$$

Once again this is a simple LHS=RHS proof using properties of the real numbers. Starting on the left we have

$$(a+b) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} (a+b)x_1 \\ (a+b)x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_1 \\ ax_2 + bx_2 \end{pmatrix}$$

$$= \begin{pmatrix} ax_1 \\ ax_2 \end{pmatrix} + \begin{pmatrix} bx_1 \\ bx_2 \end{pmatrix} = a \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} ,$$

as required.

(·iii) Additive distributivity: This time we need to check the equation The equation we need to check is

$$a \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) \stackrel{?}{=} a \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + a \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} ,$$

i.e., one scalar but two different vectors. The method is by now becoming familiar

$$a \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) = a \left(\begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix} \right) = \begin{pmatrix} a(x_1 + y_1) \\ a(x_2 + y_2) \end{pmatrix}$$

$$= \begin{pmatrix} ax_1 + ay_1 \\ ax_2 + ay_2 \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \end{pmatrix} + \begin{pmatrix} ay_1 \\ ay_2 \end{pmatrix} = a \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + a \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} ,$$

again as required.

(·iv) Multiplicative associativity. Just as for addition, this is the requirement that the order of bracketing does not matter. We need to establish whether

$$(a.b)\cdot\begin{pmatrix}x_1\\x_2\end{pmatrix} \stackrel{?}{=} a\cdot\left(b\cdot\begin{pmatrix}x_1\\x_2\end{pmatrix}\right).$$

This clearly holds for real numbers $a.(b.x)=(a.b).x$. The computation is

$$(a.b)\cdot\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}(a.b).x_1\\(a.b).x_2\end{pmatrix}=\begin{pmatrix}a.(b.x_1)\\a.(b.x_2)\end{pmatrix}=a.\begin{pmatrix}(b.x_1)\\(b.x_2)\end{pmatrix}=a\cdot\left(b\cdot\begin{pmatrix}x_1\\x_2\end{pmatrix}\right),$$

which is what we want.

(·v) Unity: We need to find a special scalar acts the way we would expect ''1'' to behave. *I.e.*

$$\text{``1''}\cdot\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}x_1\\x_2\end{pmatrix}.$$

There is an obvious choice for this special scalar---just the real number 1 itself. Indeed, to be pedantic lets calculate

$$1\cdot\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}1.x_1\\1.x_2\end{pmatrix}=\begin{pmatrix}x_1\\x_2\end{pmatrix}.$$

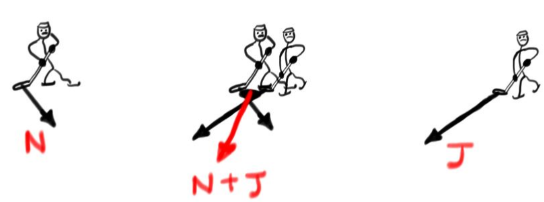
Now we are done---we have really proven the $\mathbb{R}^2$ is a vector space so lets write a little square $\square$ to celebrate.

Example of a Vector Space

This video talks about the definition of a vector space. Even though the defintion looks long, complicated and abstract, it is actually designed to model a very wide range of real life situations. As an example, consider the vector space

$$V=\{\text{all possible ways to hit a hockey puck}\}.$$

The different ways of hitting a hockey puck can all be considered as vectors. You can think about adding vectors by having two players hitting the puck at the same time. This picture shows vectors N and J corresponding to the ways Nicole Darwitz and Jenny Potter hit a hockey puck, plus the vector obtained when they hit the puck together.



You can also model the new vector $2J$ obtained by scalar multiplication by 2 by thinking about Jenny hitting the puck twice (or a world with two Jenny Potters....). Now ask yourself questions like whether the multiplicative distributive law

$$2J + 2N = 2(J + N)$$

make sense in this context.

Hint for Review Question 5

Lets worry about the last part of the problem. The problem can be solved by considering a non-zero simple polynomial, such as a degree 0 polynomial, and multiplying by $i \in \mathbb{C}$. That is to say we take a vector $p \in P_3^{\mathbb{R}}$ and then considering $i \cdot p$. This will violate one of the vector space rules about scalars, and you should take from this that the scalar field matters.

As a second hint, consider $\mathbb{Q}$ (the field of rational numbers). This is not a vector space over $\mathbb{R}$ since $\sqrt{2} \cdot 1 = \sqrt{2} \notin \mathbb{Q}$, so it is not closed under scalar multiplication, but it is clearly a vector space over $\mathbb{Q}$.

G.5 Linear Transformations

Hint for Review Question 5

The first thing we see in the problem is a definition of this new space P_n. Elements of P_n are polynomials that look like

$$a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n$$

where the a_i's are constants. So this means if L is a linear transformation from $P_2 \to P_3$ that the inputs of L are degree two polynomials which look like

$$a_0 + a_1 t + a_2 t^2$$

and the output will have degree three and look like

$$b_0 + b_1 t + b_2 t^2 + b_3 t^3$$

We also know that L is a linear transformation, so what does that mean in this case? Well, by linearity we know that we can separate out the sum, and pull out the constants so we get

$$L(a_0 + a_1 t + a_2 t^2) = a_0 L(1) + a_1 L(t) + a_2 L(t^2)$$

Just this should be really helpful for the first two parts of the problem. The third part of the problem is asking us to think about this as a linear algebra problem, so lets think about how we could write this in the vector notation we use in the class. We could write

$$a_0 + a_1 t + a_2 t^2 \text{ as } \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

And think for a second about how you add polynomials, you match up terms of the same degree and add the constants component-wise. So it makes some sense to think about polynomials this way, since vector addition is also component-wise.

We could also write the output

$$b_0 + b_1 t + b_2 t^2 + b_3 t^3 \text{ as } \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} b_3$$

Then lets look at the information given in the problem and think about it in terms of column vectors

- $L(1) = 4$ but we can think of the input $1 = 1 + 0t + 0t^2$ and the output $4 = 4 + 0t + 0t^2 0t^3$ and write this as $L(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}) = \begin{pmatrix} 4 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

- $L(t) = t^3$ This can be written as $L(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

- $L(t^2) = t - 1$ It might be a little trickier to figure out how to write $t - 1$ but if we write the polynomial out with the terms in order and with zeroes next to the terms that do not appear, we can see that

$$t - 1 = -1 + t + 0t^2 + 0t^3 \text{ corresponds to } \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

So this can be written as $L(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}) = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$

Now to think about how you would write the linear transformation L as a matrix, first think about what the dimensions of the matrix would be. Then look at the first two parts of this problem to help you figure out what the entries should be.

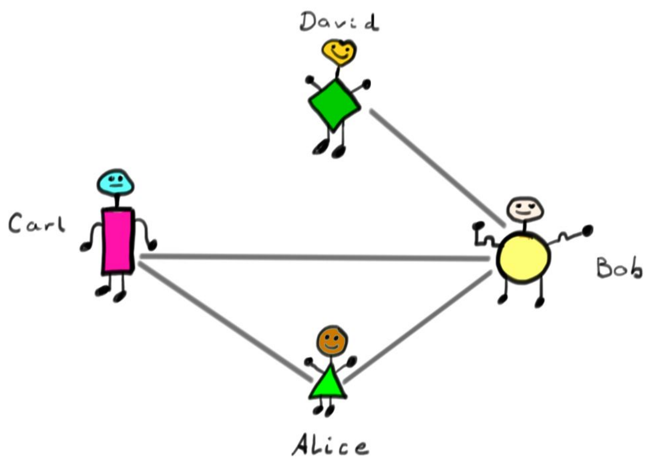
G.6 Matrices

Adjacency Matrix Example

Lets think about a graph as a mini-facebook. In this tiny facebook there are only four people, Alice, Bob, Carl, and David.

Suppose we have the following relationships

- Alice and Bob are friends.
- Alice and Carl are friends.
- Carl and Bob are friends.
- David and Bob are friends.



Now draw a picture where each person is a dot, and then draw a line between the dots of people who are friends. This is an example of a graph if you think of the people as nodes, and the friendships as edges.

Now lets make a 4×4 matrix, which is an adjacency matrix for the graph. Make a column and a row for each of the four people. It will look a lot like a table. When two people are friends put a 1 the the row of one and the column of the other. For example Alice and Carl are friends so we can label the table below.

	A	B	C	D
A			1	
B				
C	1			
D				

We can continue to label the entries for each friendship. Here lets assume that people are friends with themselves, so the diagonal will be all ones.

	A	B	C	D
A	1	1	1	0
B	1	1	1	1
C	1	1	1	0
D	0	1	0	1

Then take the entries of this table as a matrix

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

Notice that this table is symmetric across the diagonal, the same way a multiplication table would be symmetric. This is because on facebook friendship is symmetric in the sense that you can't be friends with someone if they aren't friends with you too. This is an example of a symmetric matrix.

You could think about what you would have to do differently to draw a graph for something like twitter where you don't have to follow everyone who follows you. The adjacency matrix might not be symmetric then.

Do Matrices Commute?

This video shows you a funny property of matrices. Some matrix properties look just like those for numbers. For example numbers obey

$$a(bc) = (ab)c$$

and so do matrices:

$$A(BC) = (AB)C.$$

This says the order of bracketing does not matter and is called associativity. Now we ask ourselves whether the basic property of numbers

$$ab = ba\,,$$

holds for matrices

$$AB \stackrel{?}{=} BA\,.$$

For this, firstly note that we need to work with square matrices even for both orderings to even make sense. Lets take a simple 2×2 example, let

$$A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}\,, \qquad B = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}\,, \qquad C = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}\,.$$

In fact, computing AB and BA we get the same result

$$AB = BA = \begin{pmatrix} 1 & a+b \\ 0 & 1 \end{pmatrix}\,,$$

so this pair of matrices do commute. Lets try A and C:

$$AC = \begin{pmatrix} 1+a^2 & a \\ a & 1 \end{pmatrix}, \qquad \text{and} \qquad CA = \begin{pmatrix} 1 & a \\ a & 1+a^2 \end{pmatrix}$$

so

$$AC \neq CA$$

and this pair of matrices does *not* commute. Generally, matrices usually do not commute, and the problem of finding those that do is a very interesting one.

Matrix Exponential Example

This video shows you how to compute

$$\exp \begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}.$$

For this we need to remember that the matrix exponential is defined by its power series

$$\exp M := I + M + \frac{1}{2!}M^2 + \frac{1}{3!}M^3 + \cdots .$$

Now lets call

$$\begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix} = i\theta$$

where the *matrix*

$$i := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and by matrix multiplication is seen to obey

$$i^2 = -I, \qquad i^3 = -i, i^4 = I.$$

Using these facts we compute by organizing terms according to whether they have an i or not:

$$\begin{aligned}
\exp i\theta &= I + \frac{1}{2!}\,\theta^2(-I) + \frac{1}{4!}\,(+I) + \cdots \\
&+ i\theta + \frac{1}{3!}\,\theta^3(-i) + \frac{1}{5!}\,i + \cdots \\
&= I(1 - \frac{1}{2!}\,\theta^2 + \frac{1}{4!}\,\theta^4 + \cdots) \\
&+ i(\theta - \frac{1}{3!}\,\theta^3 + \frac{1}{5!}\,\theta^5 + \cdots) \\
&= I\cos\theta + i\sin\theta \\
&= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.
\end{aligned}$$

Here we used the familiar Taylor series for the cosine and sine functions. A fun thing to think about is how the above matrix acts on vector in the plane.

Proof Explanation

In this video we will talk through the steps required to prove

$$\operatorname{tr} MN = \operatorname{tr} NM\,.$$

There are some useful things to remember, first we can write

$$M = (m^i_j) \qquad \text{and} \qquad N = (n^i_j)$$

where the upper index labels rows and the lower one columns. Then

$$MN = \big(\sum_l m^i_l n^l_j\big)\,,$$

where the ''open'' indices i and j label rows and columns, but the index l is a ''dummy'' index because it is summed over. (We could have given it any name we liked!).

Finally the trace is the sum over diagonal entries for which the row and column numbers must coincide

$$\operatorname{tr} M = \sum_i m^i_i\,.$$

Hence starting from the left of the statement we want to prove, we have

$$\texttt{LHS} = \operatorname{tr} MN = \sum_i \sum_l m^i_l n^l_i\,.$$

Next we do something obvious, just change the order of the entries m^i_l and n^l_i (they are just numbers) so

$$\sum_i \sum_l m^i_l n^l_i = \sum_i \sum_l n^l_i m^i_l\,.$$

Equally obvious, we now rename $i \to l$ and $l \to i$ so

$$\sum_i \sum_l m^i_l n^l_i = \sum_l \sum_i n^i_l m^l_i\,.$$

Finally, since we have finite sums it is legal to change the order of summations

$$\sum_l \sum_i n^i_l m^l_i = \sum_i \sum_l n^i_l m^l_i\,.$$

This expression is the same as the one on the line above where we started except the m and n have been swapped so

$$\sum_i \sum_l m^i_l n^l_i = \operatorname{tr} NM = \texttt{RHS}\,.$$

This completes the proof. □

Hint for Review Question 4

This problem just amounts to remembering that the dot product of $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ is

$$x_1y_1 + x_2y_2 + \cdots + x_ny_n\,.$$

Then try multiplying the above row vector times y^T and compare.

Hint for Review Question 5

The majority of the problem comes down to showing that matrices are right distributive. Let M_k is all $n \times k$ matrices for any n, and define the map $f_R\colon M_k \to M_m$ by $f_R(M) = MR$ where R is some $k \times m$ matrix. It should be clear that $f_R(\alpha \cdot M) = (\alpha M)R = \alpha(MR) = \alpha f_R(M)$ for any scalar α. Now all that needs to be proved is that

$$f_R(M + N) = (M + N)R = MR + NR = f_R(M) + f_R(N),$$

and you can show this by looking at each entry.

We can actually generalize the concept of this problem. Let V be some vector space and $\mathbb{M}$ be some collection of matrices, and we say that $\mathbb{M}$ is a *left-action* on V if

$$(M \cdot N) \circ v = M \circ (N \circ v)$$

for all $M, N \in \mathbb{N}$ and $v \in V$ where $\cdot$ denoted multiplication in $\mathbb{M}$ (i.e. standard matrix multiplication) and $\circ$ denotes the matrix is a linear map on a vector (i.e. $M(v)$). There is a corresponding notion of a right action where

$$v \circ (M \cdot N) = (v \circ M) \circ N$$

where we treat $v \circ M$ as $M(v)$ as before, and note the order in which the matrices are applied. People will often omit the left or right because they are essentially the same, and just say that $\mathbb{M}$ acts on V.

Hint for Review Question 8

This is a hint for computing exponents of matrices. So what is e^A if A is a matrix? We remember that the Taylor series for

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

So as matrices we can think about

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

This means we are going to have an idea of what A^n looks like for any n. Lets look at the example of one of the matrices in the problem. Let

$$A = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}.$$

Lets compute A^n for the first few n.

$$\begin{aligned} A^0 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ A^1 &= \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \\ A^2 &= A \cdot A = \begin{pmatrix} 1 & 2\lambda \\ 0 & 1 \end{pmatrix} \\ A^3 &= A^2 \cdot A = \begin{pmatrix} 1 & 3\lambda \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

There is a pattern here which is that

$$A^n = \begin{pmatrix} 1 & n\lambda \\ 0 & 1 \end{pmatrix},$$

then we can think about the first few terms of the sequence

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} = A^0 + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots.$$

Looking at the entries when we add this we get that the upper left-most entry looks like this:

$$1 + 1 + \frac{1}{2} + \frac{1}{3!} + \dots = \sum_{n=0}^{\infty} \frac{1}{n!} = e^1.$$

Continue this process with each of the entries using what you know about Taylor series expansions to find the sum of each entry.

2×2 Example

Lets go though and show how this 2×2 example satisfies all of these properties. Lets look at

$$M = \begin{pmatrix} 7 & 3 \\ 11 & 5 \end{pmatrix}$$

We have a rule to compute the inverse

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

So this means that

$$M^{-1} = \frac{1}{35-33}\begin{pmatrix} 5 & -3 \\ -11 & 7 \end{pmatrix}$$

Lets check that $M^{-1}M = I = MM^{-1}$.

$$M^{-1}M = \frac{1}{35-33}\begin{pmatrix} 5 & -3 \\ -11 & 7 \end{pmatrix}\begin{pmatrix} 7 & 3 \\ 11 & 5 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = I$$

You can compute MM^{-1}, this should work the other way too.

Now lets think about products of matrices

$$\text{Let } A = \begin{pmatrix} 1 & 3 \\ 1 & 5 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

Notice that $M = AB$. We have a rule which says that $(AB)^{-1} = B^{-1}A^{-1}$. Lets check to see if this works

$$A^{-1} = \frac{1}{2}\begin{pmatrix} 5 & -3 \\ -1 & 1 \end{pmatrix} \text{ and } B^{-1} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$$

and

$$B^{-1}A^{-1} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 5 & -3 \\ -1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Hint for Review Problem 3

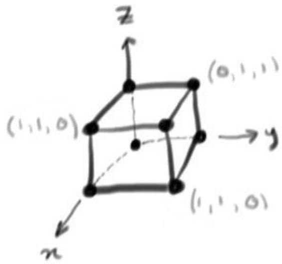
Firstnote that (b) implies (a) is the easy direction: just think about what it means for M to be non-singular and for a linear function to be well-defined. Therefore we assume that M is singular which implies that there exists a non-zero vector X_0 such that $MX_0 = 0$. Now assume there exists some vector X_V such that $MX_V = V$, and look at what happens to $X_V + c \cdot X_0$ for any c in your field. Lastly don't forget to address what happens if X_V does not exist.

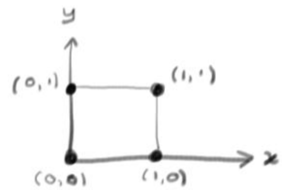
Hint for Review Question 4

In the text, only inverses for square matrices were discussed, but there is a notion of left and right inverses for matrices that are not square. It helps to look at an example with bits to see why. To start with we look at vector spaces

$$\mathbb{Z}_2^3 = \{(x,y,z) | x,y,z = 0,1\} \qquad \text{and} \qquad \mathbb{Z}_2^2 = \{(x,y) | x,y = 0,1\}\,.$$

These have 8 and 4 vectors, respectively, that can be depicted as corners of a cube or square:

$\mathbb{Z}_2^3$ 

or $\mathbb{Z}_2^2$ 

Now lets consider a linear transformation

$$L : \mathbb{Z}_2^3 \longrightarrow \mathbb{Z}_2^2 .$$

This must be represented by a matrix, and lets take the example

$$L \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} := AX .$$

Since we have bits, we can work out what L does to every vector, this is listed below

$$\begin{array}{ccc} (0,0,0) & \stackrel{L}{\mapsto} & (0,0) \\ (0,0,1) & \stackrel{L}{\mapsto} & (1,0) \\ (1,1,0) & \stackrel{L}{\mapsto} & (1,0) \\ (1,0,0) & \stackrel{L}{\mapsto} & (0,1) \\ (0,1,1) & \stackrel{L}{\mapsto} & (0,1) \\ (0,1,0) & \stackrel{L}{\mapsto} & (1,1) \\ (1,0,1) & \stackrel{L}{\mapsto} & (1,1) \\ (1,1,1) & \stackrel{L}{\mapsto} & (1,1) \end{array}$$

Now lets think about left and right inverses. A left inverse B to the matrix A would obey

$$BA = I$$

and since the identity matrix is square, B must be 2×3. It would have to undo the action of A and return vectors in $\mathbb{Z}_2^3$ to where they started from. But above, we see that different vectors in $\mathbb{Z}_2^3$ are mapped to the same vector in $\mathbb{Z}_2^2$ by the linear transformation L with matrix A. So B cannot exist. However a right inverse C obeying

$$AC = I$$

can. It would be 2×2. Its job is to take a vector in $\mathbb{Z}_2^2$ back to one in $\mathbb{Z}_2^3$ in a way that gets undone by the action of A. This can be done, but not uniquely.

Using an LU Decomposition

Lets go through how to use a LU decomposition to speed up solving a system of equations. Suppose you want to solve for x in the equation $Mx = b$

$$\begin{pmatrix} 1 & 0 & -5 \\ 3 & -1 & -14 \\ 1 & 0 & -3 \end{pmatrix} x = \begin{pmatrix} 6 \\ 19 \\ 4 \end{pmatrix}$$

where you are given the decomposition of M into the product of L and U which are lower and upper and lower triangular matrices respectively.

$$M = \begin{pmatrix} 1 & 0 & -5 \\ 3 & -1 & -14 \\ 1 & 0 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -5 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = LU$$

First you should solve $L(Ux) = b$ for Ux. The augmented matrix you would use looks like this

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 6 \\ 3 & 1 & 0 & 19 \\ 1 & 0 & 2 & 4 \end{array}\right)$$

This is an easy augmented matrix to solve because it is upper triangular. If you were to write out the three equations using variables, you would find that the first equation has already been solved, and is ready to be plugged into the second equation. This backward substitution makes solving the system much faster. Try it and in a few steps you should be able to get

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right)$$

This tells us that $Ux = \begin{pmatrix} 6 \\ 1 \\ -1 \end{pmatrix}$. Now the second part of the problem is to solve for x. The augmented matrix you get is

$$\left(\begin{array}{ccc|c} 1 & 0 & -5 & 6 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right)$$

It should take only a few step to transform it into

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -1 \end{array}\right),$$

which gives us the answer $x = \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}$.

Another LU Decomposition Example

Here we will perform an LU decomposition on the matrix

$$M = \begin{pmatrix} 1 & 7 & 2 \\ -3 & -21 & 4 \\ 1 & 6 & 3 \end{pmatrix}$$

following the procedure outlined in Section 7.7.2. So initially we have $L_1 = I_3$ and $U_1 = M$, and hence

$$L_2 = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \qquad U_2 = \begin{pmatrix} 1 & 7 & 2 \\ 0 & 0 & 10 \\ 0 & -1 & -1 \end{pmatrix}.$$

However we now have a problem since $0 \cdot c = 0$ for any value of c since we are working over a field, but we can quickly remedy this by swapping the second and third rows of U_2 to get U_2' and note that we just interchange the corresponding rows all columns left of and including the column we added values to in L_2 to get L_2'. Yet this gives us a small problem as $L_2'U_2' \neq M$; in fact it gives us the similar matrix M' with the second and third rows swapped. In our original problem $MX = V$, we also need to make the corresponding swap on our vector V to get a V' since all of this amounts to changing the order of our two equations, and note that this clearly does not change the solution. Back to our example, we have

$$L_2' = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \qquad U_2' = \begin{pmatrix} 1 & 7 & 2 \\ 0 & -1 & -1 \\ 0 & 0 & 10 \end{pmatrix},$$

and note that U_2' is upper triangular. Finally you can easily see that

$$L_2'U_2' = \begin{pmatrix} 1 & 7 & 2 \\ 1 & 6 & 3 \\ -3 & -21 & 4 \end{pmatrix} = M'$$

which solves the problem of $L_2'U_2'X = M'X = V'$. (We note that as augmented matrices $(M'|V') \sim (M|V)$.)

Block LDU Explanation

This video explains how to do a block LDU decomposition. Firstly remember some key facts about block matrices: It is important that the blocks fit together properly. For example, if we have matrices

matrix	shape
X	$r \times r$
Y	$r \times t$
Z	$t \times r$
W	$t \times t$

we could fit these together as a $(r+t)\times(r+t)$ square block matrix

$$M = \left(\begin{array}{c|c} X & Y \\ \hline Z & W \end{array}\right).$$

Matrix multiplication works for blocks just as for matrix entries:

$$M^2 = \left(\begin{array}{c|c} X & Y \\ \hline Z & W \end{array}\right)\left(\begin{array}{c|c} X & Y \\ \hline Z & W \end{array}\right) = \left(\begin{array}{c|c} X^2+YZ & XY+YW \\ \hline ZX+WZ & ZY+W^2 \end{array}\right).$$

Now lets specialize to the case where the square matrix X has an inverse. Then we can multiply out the following triple product of a lower triangular, a block diagonal and an upper triangular matrix:

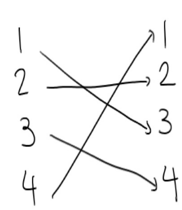
$$\begin{aligned}
&\left(\begin{array}{c|c} I & 0 \\ \hline ZX^{-1} & I \end{array}\right)\left(\begin{array}{c|c} X & 0 \\ \hline 0 & W-ZX^{-1}Y \end{array}\right)\left(\begin{array}{c|c} I & X^{-1}Y \\ \hline 0 & I \end{array}\right) \\
&= \left(\begin{array}{c|c} X & 0 \\ \hline Z & W-ZX^{-1}Y \end{array}\right)\left(\begin{array}{c|c} I & X^{-1}Y \\ \hline 0 & I \end{array}\right) \\
&= \left(\begin{array}{c|c} X & Y \\ \hline ZX^{-1}Y+Z & W-ZX^{-1}Y \end{array}\right) \\
&= \left(\begin{array}{c|c} X & Y \\ \hline Z & W \end{array}\right) = M\,.
\end{aligned}$$

This shows that the LDU decomposition given in Section 7.7 is correct.

G.7 Determinants

Permutation Example

Lets try to get the hang of permutations. A permutation is a function which scrambles things. Suppose we had



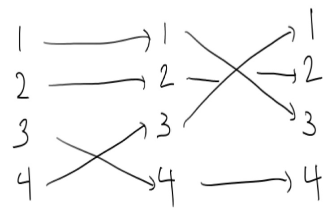
This looks like a function σ that has values

$$\sigma(1)=3,\ \sigma(2)=2,\ \sigma(3)=4,\ \sigma(4)=1\,.$$

Then we could write this as

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ \sigma(1) & \sigma(2) & \sigma(3) & \sigma(4) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{bmatrix}$$

We could write this permutation in two steps by saying that first we swap 3 and 4, and then we swap 1 and 3. The order here is important.



This is an even permutation, since the number of swaps we used is two (an even number).

Elementary Matrices

This video will explain some of the ideas behind elementary matrices. First think back to linear systems, for example n equations in n unknowns:

$$\left\{ \begin{array}{rcl} a^1_1 x^1 + a^1_2 x^2 + \cdots + a^1_n x^n &=& v^1 \\ a^2_1 x^1 + a^2_2 x^2 + \cdots + a^2_n x^n &=& v^2 \\ &\vdots& \\ a^n_1 x^1 + a^n_2 x^2 + \cdots + a^n_n x^n &=& v^n \,. \end{array} \right.$$

We know it is helpful to store the above information with matrices and vectors

$$M := \begin{pmatrix} a^1_1 & a^1_2 & \cdots & a^1_n \\ a^2_1 & a^2_2 & \cdots & a^2_n \\ \vdots & \vdots & & \vdots \\ a^n_1 & a^n_2 & \cdots & a^n_n \end{pmatrix}, \qquad X := \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix}, \qquad V := \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix}.$$

Here we will focus on the case the M is square because we are interested in its inverse M^{-1} (if it exists) and its determinant (whose job it will be to determine the existence of M^{-1}).

We know at least three ways of handling this linear system problem:

1. As an augmented matrix

   $$\left(\begin{array}{c|c} M & V \end{array} \right).$$

   Here our plan would be to perform row operations until the system looks like

   $$\left(\begin{array}{c|c} I & M^{-1}V \end{array} \right),$$

   (assuming that M^{-1} exists).

2. As a matrix equation
$$MX = V\,,$$
which we would solve by finding M^{-1} (again, if it exists), so that
$$X = M^{-1}V\,.$$

3. As a linear transformation
$$L : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$
via
$$\mathbb{R}^n \ni X \longmapsto MX \in \mathbb{R}^n\,.$$
In this case we have to study the equation $L(X) = V$ because $V \in \mathbb{R}^n$.

Lets focus on the first two methods. In particular we want to think about how the augmented matrix method can give information about finding M^{-1}. In particular, how it can be used for handling determinants.

The main idea is that the row operations changed the augmented matrices, but we also know how to change a matrix M by multiplying it by some other matrix E, so that $M \to EM$. In particular can we find ''elementary matrices'' the perform row operations?

Once we find these elementary matrices is is *very important* to ask how they effect the determinant, but you can think about that for your own self right now.

Lets tabulate our names for the matrices that perform the various row operations:

Row operation	Elementary Matrix
$R_i \leftrightarrow R_j$	E^i_j
$R_i \to \lambda R_i$	$R^i(\lambda)$
$R_i \to R_i + \lambda R_j$	$S^i_j(\lambda)$

To finish off the video, here is how all these elementary matrices work for a 2×2 example. Lets take
$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

A good thing to think about is what happens to $\det M = ad - bc$ under the operations below.

- Row swap:
$$E^1_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad E^1_2 M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix}.$$

- Scalar multiplying:

$$R^1(\lambda) = \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}, \qquad E^1_2 M = \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b \\ c & d \end{pmatrix}.$$

- Row sum:

$$S^1_2(\lambda) = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}, \quad S^1_2(\lambda) M = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + \lambda c & b + \lambda d \\ c & d \end{pmatrix}.$$

Elementary Determinants

This video will show you how to calculate determinants of elementary matrices. First remember that the job of an elementary row matrix is to perform row operations, so that if E is an elementary row matrix and M some given matrix,

$$EM$$

is the matrix M with a row operation performed on it.

The next thing to remember is that the determinant of the identity is 1. Moreover, we also know what row operations do to determinants:

- Row swap E^i_j: flips the sign of the determinant.
- Scalar multiplication $R^i(\lambda)$: multiplying a row by λ multiplies the determinant by λ.
- Row addition $S^i_j(\lambda)$: adding some amount of one row to another does not change the determinant.

The corresponding elementary matrices are obtained by performing exactly these operations on the identity:

$$E^i_j = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix},$$

$$R^i(\lambda) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix},$$

$$S^i_j(\lambda) = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & \lambda & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

So to calculate their determinants, we just have to apply the above list of what happens to the determinant of a matrix under row operations to the determinant of the identity. This yields

$$\det E^i_j = -1\,, \qquad \det R^i(\lambda) = \lambda\,, \qquad \det S^i_j(\lambda) = 1\,.$$

Determinants and Inverses

Lets figure out the relationship between determinants and invertibility. If we have a system of equations $Mx = b$ and we have the inverse M^{-1} then if we multiply on both sides we get $x = M^{-1}Mx = M^{-1}b$. If the inverse exists we can solve for x and get a solution that looks like a point.

So what could go wrong when we want solve a system of equations and get a solution that looks like a point? Something would go wrong if we didn't have enough equations for example if we were just given

$$x + y = 1$$

or maybe, to make this a square matrix M we could write this as

$$\begin{aligned} x + y &= 1 \\ 0 &= 0 \end{aligned}$$

The matrix for this would be $M = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and det$(M) = 0$. When we compute the determinant, this row of all zeros gets multiplied in every term. If instead we were given redundant equations

$$\begin{aligned} x + y &= 1 \\ 2x + 2y &= 2 \end{aligned}$$

The matrix for this would be $M = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ and det$(M) = 0$. But we know that with an elementary row operation, we could replace the second row with a row

of all zeros. Somehow the determinant is able to detect that there is only one equation here. Even if we had a set of contradictory set of equations such as

$$\begin{aligned} x + y &= 1 \\ 2x + 2y &= 0, \end{aligned}$$

where it is not possible for both of these equations to be true, the matrix M is still the same, and still has a determinant zero.

Lets look at a three by three example, where the third equation is the sum of the first two equations.

$$\begin{aligned} x + y + z &= 1 \\ y + z &= 1 \\ x + 2y + 2z &= 2 \end{aligned}$$

and the matrix for this is

$$M = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix}$$

If we were trying to find the inverse to this matrix using elementary matrices

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 \end{array}\right) = \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right)$$

And we would be stuck here. The last row of all zeros cannot be converted into the bottom row of a 3×3 identity matrix. this matrix has no inverse, and the row of all zeros ensures that the determinant will be zero. It can be difficult to see when one of the rows of a matrix is a linear combination of the others, and what makes the determinant a useful tool is that with this reasonably simple computation we can find out if the matrix is invertible, and if the system will have a solution of a single point or column vector.

Alternative Proof

Here we will prove more directly that the determinant of a product of matrices is the product of their determinants. First we reference that for a matrix M with rows r_i, if M' is the matrix with rows $r'_j = r_j + \lambda r_i$ for $j \neq i$ and $r'_i = r_i$, then $\det(M) = \det(M')$ Essentially we have M' as M multiplied by the elementary row sum matrices $S^i_j(\lambda)$. Hence we can create an upper-triangular matrix U such that $\det(M) = \det(U)$ by first using the first row to set $m^1_i \mapsto 0$ for all $i > 1$, then iteratively (increasing k by 1 each time) for fixed k using the k-th row to set $m^k_i \mapsto 0$ for all $i > k$.

Now note that for two upper-triangular matrices $U = (u_i^j)$ and $U' = (u_i'^j)$, by matrix multiplication we have $X = UU' = (x_i^j)$ is upper-triangular and $x_i^i = u_i^i u_i'^i$. Also since every permutation would contain a lower diagonal entry (which is 0) have $\det(U) = \prod_i u_i^i$. Let A and A' have corresponding upper-triangular matrices U and U' respectively (i.e. $\det(A) = \det(U)$), we note that AA' has a corresponding upper-triangular matrix UU', and hence we have

$$\begin{aligned}\det(AA') = \det(UU') &= \prod_i u_i^i u_i'^i \\ &= \left(\prod_i u_i^i\right)\left(\prod_i u_i'^i\right) \\ &= \det(U)\det(U') = \det(A)\det(A').\end{aligned}$$

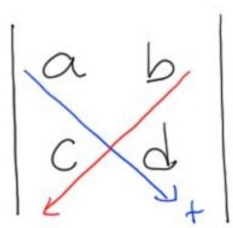
Practice taking Determinants

Lets practice taking determinants of 2×2 and 3×3 matrices.

For 2×2 matrices we have a formula

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc\,.$$

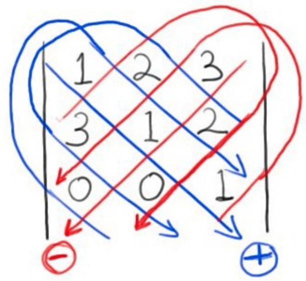
This formula might be easier to remember if you think about this picture.



Now we can look at three by three matrices and see a few ways to compute the determinant. We have a similar pattern for 3×3 matrices. Consider the example

$$\det \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = ((1 \cdot 1 \cdot 1) + (2 \cdot 2 \cdot 0) + (3 \cdot 3 \cdot 0)) - ((3 \cdot 1 \cdot 0) + (1 \cdot 2 \cdot 0) + (3 \cdot 2 \cdot 1)) = -5$$

We can draw a picture with similar diagonals to find the terms that will be positive and the terms that will be negative.



Another way to compute the determinant of a matrix is to use this recursive formula. Here I take the coefficients of the first row and multiply them by the determinant of the minors and the cofactor. Then we can use the formula for a two by two determinant to compute the determinant of the minors

$$\text{det}\begin{pmatrix}1&2&3\\3&1&2\\0&0&1\end{pmatrix}=1\begin{vmatrix}1&2\\0&1\end{vmatrix}-2\begin{vmatrix}3&2\\0&1\end{vmatrix}+3\begin{vmatrix}3&1\\0&0\end{vmatrix}=1(1-0)-2(3-0)+3(0-0)=-5$$

Decide which way you prefer and get good at taking determinants, you'll need to compute them in a lot of problems.

Hint for Review Problem 5

For an arbitrary 3×3 matrix $A=(a^i_j)$, we have

$$\det(A)=a^1_1a^2_2a^3_3+a^1_2a^2_3a^3_1+a^1_3a^2_1a^3_2-a^1_1a^2_3a^3_2-a^1_2a^2_1a^3_3-a^1_3a^2_2a^3_1$$

and so the complexity is $5a+12m$. Now note that in general, the complexity c_n of the expansion minors formula of an arbitrary $n\times n$ matrix should be

$$c_n=(n-1)a+nc_{n-1}m$$

since $\det(A)=\sum_{i=1}^{n}(-1)^i a^1_i\,\text{cofactor}(a^1_i)$ and $\text{cofactor}(a^1_i)$ is an $(n-1)\times(n-1)$ matrix. This is one way to prove part (c).

G.8 Subspaces and Spanning Sets

Linear systems as spanning sets

Suppose that we were given a set of linear equations $l^j(x^1, x^2, \dots, x^n)$ and we want to find out if $l^j(X) = v^j$ for all j for some vector $V = (v^j)$. We know that we can express this as the matrix equation

$$\sum_i l_i^j x^i = v^j$$

where l_i^j is the coefficient of the variable x^i in the equation l^j. However, this is also stating that V is in the span of the vectors $\{L_i\}_i$ where $L_i = (l_i^j)_j$. For example, consider the set of equations

$$\begin{aligned} 2x + 3y - z &= 5 \\ -x + 3y + z &= 1 \\ x + y - 2z &= 3 \end{aligned}$$

which corresponds to the matrix equation

$$\begin{pmatrix} 2 & 3 & -1 \\ -1 & 3 & 1 \\ 1 & 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \\ 3 \end{pmatrix}.$$

We can thus express this problem as determining if the vector

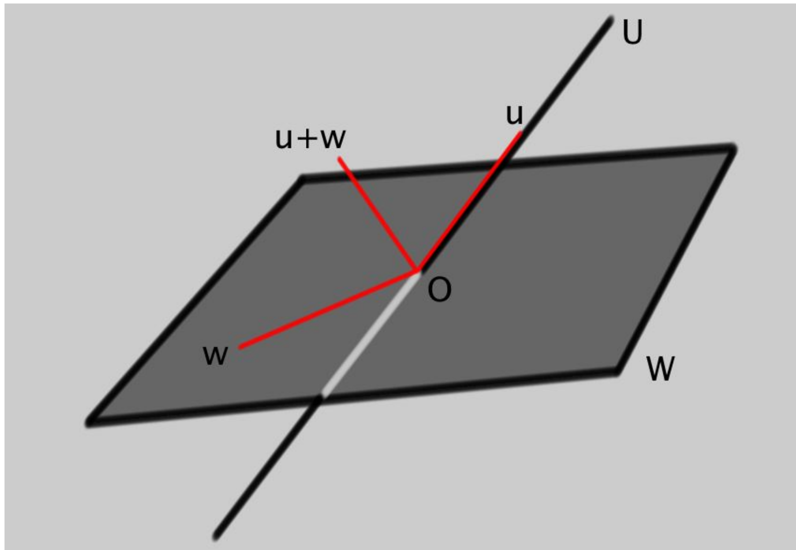
$$V = \begin{pmatrix} 5 \\ 1 \\ 3 \end{pmatrix}$$

lies in the span of

$$\left\{ \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -2 \end{pmatrix} \right\}.$$

Hint for Review Problem 2

For the first part, try drawing an example in $\mathbb{R}^3$:



Here we have taken the subspace W to be a plane through the origin and U to be a line through the origin. The hint now is to think about what happens when you add a vector $u \in U$ to a vector $w \in W$. Does this live in the union $U \cup W$?

For the second part, we take a more theoretical approach. Lets suppose that $v \in U \cap W$ and $v' \in U \cap W$. This implies

$$v \in U \quad \text{and} \quad v' \in U\,.$$

So, since U is a subspace and all subspaces are vector spaces, we know that the linear combination

$$\alpha v + \beta v' \in U\,.$$

Now repeat the same logic for W and you will be nearly done.

G.9 Linear Independence

Worked Example

This video gives some more details behind the example for the following four vectors in $\mathbb{R}^3$ Consider the following vectors in $\mathbb{R}^3$:

$$v_1 = \begin{pmatrix} 4\\ -1\\ 3 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} -3\\ 7\\ 4 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 5\\ 12\\ 17 \end{pmatrix}, \qquad v_4 = \begin{pmatrix} -1\\ 1\\ 0 \end{pmatrix}.$$

The example asks whether they are linearly independent, and the answer is immediate: NO, four vectors can never be linearly independent in $\mathbb{R}^3$. This vector space is simply not big enough for that, but you need to understand the

notion of the dimension of a vector space to see why. So we think the vectors v_1, v_2, v_3 and v_4 are linearly dependent, which means we need to show that there is a solution to

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 + \alpha_4 v_4 = 0$$

for the numbers α_1, α_2, α_3 and α_4 not all vanishing.

To find this solution we need to set up a linear system. Writing out the above linear combination gives

$$\begin{array}{rrrrcl} 4\alpha_1 & -3\alpha_2 & +5\alpha_3 & -\alpha_4 & = & 0\,, \\ -\alpha_1 & +7\alpha_2 & +12\alpha_3 & +\alpha_4 & = & 0\,, \\ 3\alpha_1 & +4\alpha_2 & +17\alpha_3 & & = & 0\,. \end{array}$$

This can be easily handled using an augmented matrix whose columns are just the vectors we started with

$$\left(\begin{array}{cccc|c} 4 & -3 & 5 & -1 & 0\,, \\ -1 & 7 & 12 & 1 & 0\,, \\ 3 & 4 & 17 & 0 & 0\,. \end{array}\right).$$

Since there are only zeros on the right hand column, we can drop it. Now we perform row operations to achieve RREF

$$\begin{pmatrix} 4 & -3 & 5 & -1 \\ -1 & 7 & 12 & 1 \\ 3 & 4 & 17 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & \frac{71}{25} & -\frac{4}{25} \\ 0 & 1 & \frac{53}{25} & \frac{3}{25} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This says that α_3 and α_4 are not pivot variable so are arbitrary, we set them to μ and ν, respectively. Thus

$$\alpha_1 = \Big(-\frac{71}{25}\,\mu + \frac{4}{25}\,\nu\Big)\,, \qquad \alpha_2 = \Big(-\frac{53}{25}\,\mu - \frac{3}{25}\,\nu\Big)\,, \qquad \alpha_3 = \mu\,, \qquad \alpha_4 = \nu\,.$$

Thus we have found a relationship among our four vectors

$$\Big(-\frac{71}{25}\,\mu + \frac{4}{25}\,\nu\Big)\,v_1 + \Big(-\frac{53}{25}\,\mu - \frac{3}{25}\,\nu\Big)\,v_2 + \mu\,v_3 + \mu_4\,v_4 = 0\,.$$

In fact this is not just one relation, but infinitely many, for any choice of μ, ν. The relationship quoted in the notes is just one of those choices.

Finally, since the vectors v_1, v_2, v_3 and v_4 are linearly dependent, we can try to eliminate some of them. The pattern here is to keep the vectors that correspond to columns with pivots. For example, setting $\mu = -1$ (say) and $\nu = 0$ in the above allows us to solve for v_3 while $\mu = 0$ and $\nu = -1$ (say) gives v_4, explicitly we get

$$v_3 = \frac{71}{25}\,v_1 + \frac{53}{25}\,v_2\,, \qquad v_4 = -\frac{4}{25}\,v_3 + \frac{3}{25}\,v_4\,.$$

This eliminates v_3 and v_4 and leaves a pair of linearly independent vectors v_1 and v_2.

Worked Proof

Here we will work through a quick version of the proof of Theorem 10.1.1. Let $\{v_i\}$ denote a set of linearly dependent vectors, so $\sum_i c^i v_i = 0$ where there exists some $c^k \neq 0$. Now without loss of generality we order our vectors such that $c^1 \neq 0$, and we can do so since addition is commutative (i.e. $a+b=b+a$). Therefore we have

$$c^1 v_1 = -\sum_{i=2}^{n} c^i v_i$$
$$v_1 = -\sum_{i=2}^{n} \frac{c^i}{c^1} v_i$$

and we note that this argument is completely reversible since every $c^i \neq 0$ is invertible and $0/c^i = 0$.

Hint for Review Problem 1

Lets first remember how $\mathbb{Z}_2$ works. The only two elements are 1 and 0. Which means when you add $1+1$ you get 0. It also means when you have a vector $\vec{v} \in B^n$ and you want to multiply it by a scalar, your only choices are 1 and 0. This is kind of neat because it means that the possibilities are finite, so we can look at an entire vector space.

Now lets think about B^3 there is choice you have to make for each co-ordinate, you can either put a 1 or a 0, there are three places where you have to make a decision between two things. This means that you have $2^3 = 8$ possibilities for vectors in B^3.

When you want to think about finding a set S that will span B^3 and is linearly independent, you want to think about how many vectors you need. You will need you have enough so that you can make every vector in B^3 using linear combinations of elements in S but you don't want too many so that some of them are linear combinations of each other. I suggest trying something really simple perhaps something that looks like the columns of the identity matrix

For part (c) you have to show that you can write every one of the elements as a linear combination of the elements in S, this will check to make sure S actually spans B^3.

For part (d) if you have two vectors that you think will span the space, you can prove that they do by repeating what you did in part (c), check that every vector can be written using only copies of of these two vectors. If you don't think it will work you should show why, perhaps using an argument that counts the number of possible vectors in the span of two vectors.

G.10 Basis and Dimension

Proof Explanation

Lets walk through the proof of theorem 11.0.1. We want to show that for $S = \{v_1, \ldots, v_n\}$ a basis for a vector space V, then every vector $w \in V$ can be written *uniquely* as a linear combination of vectors in the basis S:

$$w = c^1 v_1 + \cdots + c^n v_n.$$

We should remember that since S is a basis for V, we know two things

- $V = \operatorname{span} S$
- $v_1, \ldots, v_n$ are linearly independent, which means that whenever we have $a^1 v_1 + \ldots + a^n v_n = 0$ this implies that $a^i = 0$ for all $i = 1, \ldots, n$.

This first fact makes it easy to say that there exist constants c^i such that $w = c^1 v_1 + \cdots + c^n v_n$. What we don't yet know is that these $c^1, \ldots c^n$ are unique.

In order to show that these are unique, we will suppose that they are not, and show that this causes a contradiction. So suppose there exists a second set of constants d^i such that

$$w = d^1 v_1 + \cdots + d^n v_n\,.$$

For this to be a contradiction we need to have $c^i \neq d^i$ for some i. Then look what happens when we take the difference of these two versions of w:

$$\begin{aligned} 0_V &= w - w \\ &= (c^1 v_1 + \cdots + c^n v_n) - (d^1 v_1 + \cdots + d^n v_n) \\ &= (c^1 - d^1) v_1 + \cdots + (c^n - d^n) v_n. \end{aligned}$$

Since the v_i's are linearly independent this implies that $c^i - d^i = 0$ for all i, this means that we cannot have $c^i \neq d^i$, which is a contradiction.

Worked Example

In this video we will work through an example of how to extend a set of linearly independent vectors to a basis. For fun, we will take the vector space

$$V = \{(x, y, z, w) | x, y, z, w \in \mathbb{Z}^5\}\,.$$

This is like four dimensional space $\mathbb{R}^4$ except that the numbers can only be $\{0, 1, 2, 3, 4\}$. This is like bits, but now the rule is

$$0 = 5\,.$$

Thus, for example, $\frac{1}{4} = 4$ because $4 = 16 = 1 + 3 \times 5 = 1$. Don't get too caught up on this aspect, its a choice of base field designed to make computations go quicker!

Now, here's the problem we will solve:

$$\textbf{Find a basis for } V \textbf{ that includes the vectors } \begin{pmatrix}1\\2\\3\\4\end{pmatrix} \textbf{ and } \begin{pmatrix}0\\3\\2\\1\end{pmatrix}\textbf{.}$$

The way to proceed is to add a known (and preferably simple) basis to the vectors given, thus we consider

$$v_1 = \begin{pmatrix}1\\2\\3\\4\end{pmatrix},\ v_2 = \begin{pmatrix}0\\3\\2\\1\end{pmatrix},\ e_1 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix},\ e_2 = \begin{pmatrix}0\\1\\0\\0\end{pmatrix},\ e_3 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix},\ e_4 = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}.$$

The last four vectors are clearly a basis (make sure you understand this....) and are called the *canonical basis*. We want to keep v_1 and v_2 but find a way to turf out two of the vectors in the canonical basis leaving us a basis of four vectors. To do that, we have to study linear independence, or in other words a linear system problem defined by

$$0 = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 v_1 + \alpha_4 v_2 + \alpha_5 e_3 + \alpha_6 e_4 \,.$$

We want to find solutions for the $\alpha's$ which allow us to determine two of the $e's$. For that we use an augmented matrix

$$\left(\begin{array}{cccccc|c} 1 & 0 & 1 & 0 & 0 & 0 & 0\\ 2 & 3 & 0 & 1 & 0 & 0 & 0\\ 3 & 2 & 0 & 0 & 1 & 0 & 0\\ 4 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}\right).$$

Next comes a bunch of row operations. Note that we have dropped the last column of zeros since it has no information--you can fill in the row operations used above the $\sim$'s as an exercise:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0\\ 2 & 3 & 0 & 1 & 0 & 0\\ 3 & 2 & 0 & 0 & 1 & 0\\ 4 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0\\ 0 & 3 & 3 & 1 & 0 & 0\\ 0 & 2 & 2 & 0 & 1 & 0\\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0\\ 0 & 1 & 1 & 2 & 0 & 0\\ 0 & 2 & 2 & 0 & 1 & 0\\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0\\ 0 & 1 & 1 & 2 & 0 & 0\\ 0 & 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 3 & 0 & 1 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{pmatrix}$$

$$\sim \begin{pmatrix} \underline{1} & 0 & 1 & 0 & 0 & 0 \\ 0 & \underline{1} & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & \underline{1} & 0 & 2 \\ 0 & 0 & 0 & 0 & \underline{1} & 3 \end{pmatrix}$$

The pivots are underlined. The columns corresponding to non-pivot variables are the ones that can be eliminated--their coefficients (the α's) will be arbitrary, so set them all to zero save for the one next to the vector you are solving for which can be taken to be unity. Thus that vector can certainly be expressed in terms of previous ones. Hence, altogether, our basis is

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

Finally, as a check, note that $e_1 = v_1 + v_2$ which explains why we had to throw it away.

Hint for Review Problem 2

Since there are two possible values for each entry, we have $|B^n| = 2^n$. We note that $\dim B^n = n$ as well. Explicitly we have $B^1 = \{(0), (1)\}$ so there is only 1 basis for B^1. Similarly we have

$$B^2 = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

and so choosing any two non-zero vectors will form a basis. Now in general we note that we can build up a basis $\{e_i\}$ by arbitrarily (independently) choosing the first $i-1$ entries, then setting the i-th entry to 1 and all higher entries to 0.

G.11 Eigenvalues and Eigenvectors

2×2 Example

Here is an example of how to find the eigenvalues and eigenvectors of a 2×2 matrix.

$$M = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}.$$

Remember that an eigenvector v with eigenvalue λ for M will be a vector such that $Mv = \lambda v$ i.e. $M(v) - \lambda I(v) = \vec{0}$. When we are talking about a nonzero v then this means that $\det(M - \lambda I) = 0$. We will start by finding the eigenvalues that make this statement true. First we compute

$$\det(M - \lambda I) = \det\left(\begin{pmatrix}4 & 2\\ 1 & 3\end{pmatrix} - \begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}\right) = \det\begin{pmatrix}4-\lambda & 2\\ 1 & 3-\lambda\end{pmatrix}$$

so $\det(M - \lambda I) = (4-\lambda)(3-\lambda) - 2\cdot 1$. We set this equal to zero to find values of λ that make this true:

$$(4-\lambda)(3-\lambda) - 2\cdot 1 = 10 - 7\lambda + \lambda^2 = (2-\lambda)(5-\lambda) = 0\,.$$

This means that $\lambda = 2$ and $\lambda = 5$ are solutions. Now if we want to find the eigenvectors that correspond to these values we look at vectors v such that

$$\begin{pmatrix}4-\lambda & 2\\ 1 & 3-\lambda\end{pmatrix} v = \vec{0}\,.$$

For $\lambda = 5$

$$\begin{pmatrix}4-5 & 2\\ 1 & 3-5\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}-1 & 2\\ 1 & -2\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \vec{0}\,.$$

This gives us the equalities $-x + 2y = 0$ and $x - 2y = 0$ which both give the line $y = \frac{1}{2}x$. Any point on this line, so for example $\begin{pmatrix}2\\ 1\end{pmatrix}$, is an eigenvector with eigenvalue $\lambda = 5$.

Now lets find the eigenvector for $\lambda = 2$

$$\begin{pmatrix}4-2 & 2\\ 1 & 3-2\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}2 & 2\\ 1 & 1\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \vec{0},$$

which gives the equalities $2x + 2y = 0$ and $x + y = 0$. (Notice that these equations are not independent of one another, so our eigenvalue must be correct.) This means any vector $v = \begin{pmatrix}x\\ y\end{pmatrix}$ where $y = -x$, such as $\begin{pmatrix}1\\ -1\end{pmatrix}$, or any scalar multiple of this vector , *i.e.* any vector on the line $y = -x$ is an eigenvector with eigenvalue 2. This solution could be written neatly as

$$\lambda_1 = 5,\, v_1 = \begin{pmatrix}2\\ 1\end{pmatrix} \text{ and } \lambda_2 = 2,\, v_2 = \begin{pmatrix}1\\ -1\end{pmatrix}.$$

Jordan Block Example

Consider the matrix

$$J_2 = \begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix},$$

and we note that we can just read off the eigenvector e_1 with eigenvalue λ. However the characteristic polynomial of J_2 is $P_{J_2}(\mu) = (\mu - \lambda)^2$ so the only possible eigenvalue is λ, but we claim it does not have a second eigenvector v. To see this, we require that

$$\begin{aligned} \lambda v^1 + v^2 &= \lambda v^1 \\ \lambda v^2 &= \lambda v^2 \end{aligned}$$

which clearly implies that $v^2 = 0$. This is known as a Jordan 2-cell, and in general, a Jordan n-cell with eigenvalue λ is (similar to) the $n \times n$ matrix

$$J_n = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda & 1 \\ 0 & \cdots & 0 & 0 & \lambda \end{pmatrix}$$

which has a single eigenvector e_1.

Now consider the following matrix

$$M = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

and we see that $P_M(\lambda) = (\lambda - 3)^2(\lambda - 2)$. Therefore for $\lambda = 3$ we need to find the solutions to $(M - 3I_3)v = 0$ or in equation form:

$$\begin{aligned} v^2 &= 0 \\ v^3 &= 0 \\ -v^3 &= 0, \end{aligned}$$

and we immediately see that we must have $V = e_1$. Next for $\lambda = 2$, we need to solve $(M - 2I_3)v = 0$ or

$$\begin{aligned} v^1 + v^2 &= 0 \\ v^2 + v^3 &= 0 \\ 0 &= 0, \end{aligned}$$

and thus we choose $v^1 = 1$, which implies $v^2 = -1$ and $v^3 = 1$. Hence this is the only other eigenvector for M.

This is a specific case of Problem 13.7.

Eigenvalues

Eigenvalues and eigenvectors are extremely important. In this video we review the theory of eigenvalues. Consider a linear transformation

$$L : V \longrightarrow V$$

where $\dim V = n < \infty$. Since V is finite dimensional, we can represent L by a square matrix M by choosing a basis for V.

So the eigenvalue equation

$$\boxed{Lv = \lambda v}$$

becomes

$$\boxed{Mv = \lambda v\,,}$$

where v is a column vector and M is an $n\times n$ matrix (both expressed in whatever basis we chose for V). The scalar λ is called an eigenvalue of M and the job of this video is to show you how to find all the eigenvalues of M.

The first step is to put all terms on the left hand side of the equation, this gives

$$(M - \lambda I)v = 0\,.$$

Notice how we used the identity matrix I in order to get a matrix times v equaling zero. Now here comes a VERY important fact

$$\boxed{Nu = 0 \text{ and } u \neq 0 \iff \det N = 0\,.}$$

I.e., a square matrix can have an eigenvector with vanishing eigenvalue if and only if its determinant vanishes! Hence

$$\boxed{\det(M - \lambda I) = 0\,.}$$

The quantity on the left (up to a possible minus sign) equals the so-called characteristic polynomial

$$P_M(\lambda) := \det(\lambda I - M)\,.$$

It is a polynomial of degree n in the variable λ. To see why, try a simple 2×2 example

$$\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) = \det\begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} = (a-\lambda)(d-\lambda) - bc\,,$$

which is clearly a polynomial of order 2 in λ. For the $n\times n$ case, the order n term comes from the product of diagonal matrix elements also.

There is an amazing fact about polynomials called the *fundamental theorem of algebra*: they can always be factored over complex numbers. This means that

degree n polynomials have n complex roots (counted with multiplicity). The word can does not mean that explicit formulas for this are known (in fact explicit formulas can only be give for degree four or less). The necessity for complex numbers is easily seems from a polynomial like

$$z^2 + 1$$

whose roots would require us to solve $z^2 = -1$ which is impossible for real number z. However, introducing the imaginary unit i with

$$i^2 = -1\,,$$

we have

$$z^2 + 1 = (z - i)(z + i)\,.$$

Returning to our characteristic polynomial, we call on the fundamental theorem of algebra to write

$$P_M(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)\,.$$

The roots λ_1, λ_2,...,λ_n are the eigenvalues of M (or its underlying linear transformation L).

Eigenspaces

Consider the linear map

$$L = \begin{pmatrix} -4 & 6 & 6 \\ 0 & 2 & 0 \\ -3 & 3 & 5 \end{pmatrix}.$$

Direct computation will show that we have

$$L = Q \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} Q^{-1}$$

where

$$Q = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Therefore the vectors

$$v_1^{(2)} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \qquad v_2^{(2)} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

span the eigenspace $E^{(2)}$ of the eigenvalue 2, and for an explicit example, if we take

$$v = 2v_1^{(2)} - v_2^{(2)} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$$

we have

$$Lv = \begin{pmatrix} 2 \\ -2 \\ 4 \end{pmatrix} = 2v$$

so $v \in E^{(2)}$. In general, we note the linearly independent vectors $v_i^{(\lambda)}$ with the same eigenvalue λ span an eigenspace since for any $v = \sum_i c^i v_i^{(\lambda)}$, we have

$$Lv = \sum_i c^i L v_i^{(\lambda)} = \sum_i c^i \lambda v_i^{(\lambda)} = \lambda \sum_i c^i v_i^{(\lambda)} = \lambda v.$$

Hint for Review Problem 9

We are looking at the matrix M, and a sequence of vectors starting with $v(0) = \begin{pmatrix} x(0) \\ y(0) \end{pmatrix}$ and defined recursively so that

$$v(1) = \begin{pmatrix} x(1) \\ y(1) \end{pmatrix} = M \begin{pmatrix} x(0) \\ y(0) \end{pmatrix}.$$

We first examine the eigenvectors and eigenvalues of

$$M = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}.$$

We can find the eigenvalues and vectors by solving

$$\det(M - \lambda I) = 0$$

for λ.

$$\det \begin{pmatrix} 3-\lambda & 2 \\ 2 & 3-\lambda \end{pmatrix} = 0$$

By computing the determinant and solving for λ we can find the eigenvalues $\lambda = 1$ and 5, and the corresponding eigenvectors. You should do the computations to find these for yourself.

When we think about the question in part (b) which asks to find a vector $v(0)$ such that $v(0) = v(1) = v(2) \ldots$, we must look for a vector that satisfies $v = Mv$. What eigenvalue does this correspond to? If you found a $v(0)$ with this property would $cv(0)$ for a scalar c also work? Remember that eigenvectors have to be nonzero, so what if $c = 0$?

For part (c) if we tried an eigenvector would we have restrictions on what the eigenvalue should be? Think about what it means to be pointed in the same direction.

G.12 Diagonalization

Non Diagonalizable Example

First recall that the derivative operator is linear and that we can write it as the matrix

$$\frac{d}{dx} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 2 & 0 & \cdots \\ 0 & 0 & 0 & 3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

We note that this transforms into an infinite Jordan cell with eigenvalue 0 or

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

which is in the basis $\{n^{-1}x^n\}_n$ (where for $n = 0$, we just have 1). Therefore we note that 1 (constant polynomials) is the only eigenvector with eigenvalue 0 for polynomials since they have finite degree, and so the derivative is not diagonalizable. Note that we are ignoring infinite cases for simplicity, but if you want to consider infinite terms such as convergent series or all formal power series where there is no conditions on convergence, there are many eigenvectors. Can you find some? This is an example of how things can change in infinite dimensional spaces.

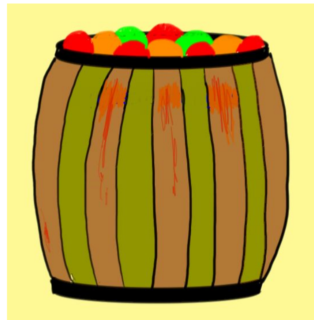
For a more finite example, consider the space $\mathbb{P}_3^{\mathbb{C}}$ of complex polynomials of degree at most 3, and recall that the derivative D can be written as

$$D = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

You can easily check that the only eigenvector is 1 with eigenvalue 0 since D always lowers the degree of a polynomial by 1 each time it is applied. Note that this is a nilpotent matrix since $D^4 = 0$, but the only nilpotent matrix that is ‘‘diagonalizable’’ is the 0 matrix.

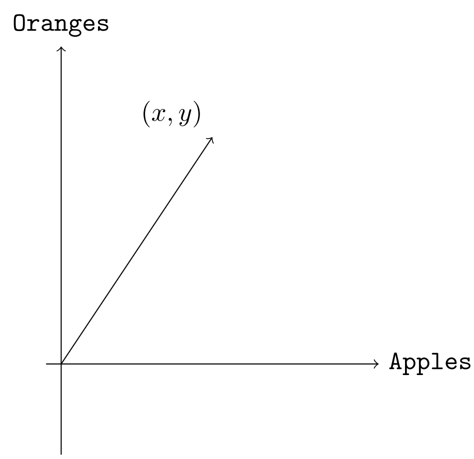
Change of Basis Example

This video returns to the example of a barrel filled with fruit



as a demonstration of changing basis.

Since this was a linear systems problem, we can try to represent what's in the barrel using a vector space. The first representation was the one where $(x, y) = (\text{apples}, \text{oranges})$:

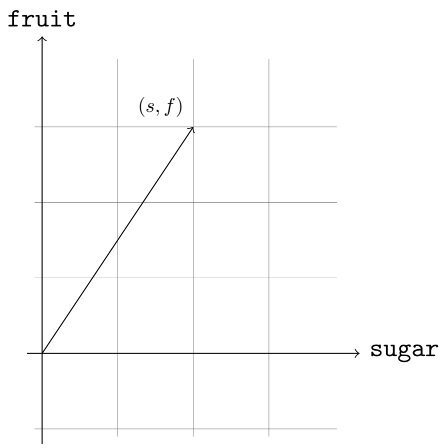


Calling the basis vectors $\vec{e}_1 := (1, 0)$ and $\vec{e}_2 := (0, 1)$, this representation would label what's in the barrel by a vector

$$\vec{x} := x\vec{e}_1 + y\vec{e}_2 = \begin{pmatrix} \vec{e}_1 & \vec{e}_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Since this is the method ordinary people would use, we will call this the ''engineer's'' method!

But this is not the approach nutritionists would use. They would note the amount of sugar and total number of fruit (s, f):

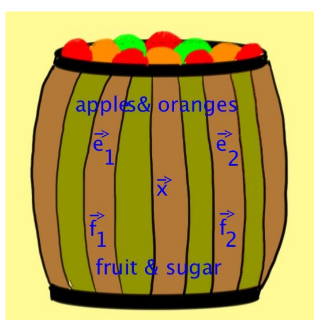


WARNING: To make sense of what comes next you need to allow for the possibity of a negative amount of fruit or sugar. This would be just like a bank, where if money is owed to somebody else, we can use a minus sign.

The vector $\vec{x}$ says what is in the barrel and does not depend which mathematical description is employed. The way nutritionists label $\vec{x}$ is in terms of a pair of basis vectors $\vec{f}_1$ and $\vec{f}_2$:

$$\vec{x} = s\vec{f}_1 + f\vec{f}_2 = \begin{pmatrix} \vec{f}_1 & \vec{f}_2 \end{pmatrix} \begin{pmatrix} s \\ f \end{pmatrix} .$$

Thus our vector space now has a bunch of interesting vectors:



The vector $\vec{x}$ labels generally the contents of the barrel. The vector $\vec{e}_1$ corresponds to one apple and one orange. The vector $\vec{e}_2$ is one orange and no apples. The vector $\vec{f}_1$ means one unit of sugar and zero total fruit (to achieve this you could lend out some apples and keep a few oranges). Finally the vector $\vec{f}_2$ represents a total of one piece of fruit and no sugar.

You might remember that the amount of sugar in an apple is called λ while oranges have twice as much sugar as apples. Thus

$$\left\{ \begin{array}{l} s = \lambda\,(x + 2y) \\ f = x + y\,. \end{array} \right.$$

Essentially, this is already our change of basis formula, but lets play around and put it in our notations. First we can write this as a matrix

$$\begin{pmatrix} s \\ f \end{pmatrix} = \begin{pmatrix} \lambda & 2\lambda \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} .$$

We can easily invert this to get

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -\frac{1}{\lambda} & 2 \\ \frac{1}{\lambda} & -1 \end{pmatrix} \begin{pmatrix} s \\ f \end{pmatrix} .$$

Putting this in the engineer's formula for $\vec{x}$ gives

$$\vec{x} = \begin{pmatrix} \vec{e}_1 & \vec{e}_2 \end{pmatrix} \begin{pmatrix} -\frac{1}{\lambda} & 2 \\ \frac{1}{\lambda} & -1 \end{pmatrix} \begin{pmatrix} s \\ f \end{pmatrix} = \begin{pmatrix} -\frac{1}{\lambda}(\vec{e}_1 - \vec{e}_2) & 2\vec{e}_1 - 2\vec{e}_2 \end{pmatrix} \begin{pmatrix} s \\ f \end{pmatrix} .$$

Comparing to the nutritionist's formula for the same object $\vec{x}$ we learn that

$$\vec{f}_1 = -\frac{1}{\lambda}\big(\vec{e}_1 - \vec{e}_2\big) \quad \text{and} \quad \vec{f}_2 = 2\vec{e}_1 - 2\vec{e}_2 .$$

Rearranging these equation we find the change of base matrix P from the engineer's basis to the nutritionist's basis:

$$\begin{pmatrix} \vec{f}_1 & \vec{f}_2 \end{pmatrix} = \begin{pmatrix} \vec{e}_1 & \vec{e}_2 \end{pmatrix} \begin{pmatrix} -\frac{1}{\lambda} & 2 \\ \frac{1}{\lambda} & -1 \end{pmatrix} =: \begin{pmatrix} \vec{e}_1 & \vec{e}_2 \end{pmatrix} P .$$

We can also go the other direction, changing from the nutritionist's basis to the engineer's basis

$$\begin{pmatrix} \vec{e}_1 & \vec{e}_2 \end{pmatrix} = \begin{pmatrix} \vec{f}_1 & \vec{f}_2 \end{pmatrix} \begin{pmatrix} \lambda & 2\lambda \\ 1 & 1 \end{pmatrix} =: \begin{pmatrix} \vec{f}_1 & \vec{f}_2 \end{pmatrix} Q .$$

Of course, we must have

$$Q = P^{-1} ,$$

(which is in fact how we constructed P in the first place).

Finally, lets consider the very first linear systems problem, where you were given that there were 27 pieces of fruit in total and twice as many oranges as apples. In equations this says just

$$x + y = 27 \quad \text{and} \quad 2x - y = 0 .$$

But we can also write this as a matrix system

$$MX = V$$

where

$$M := \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} , \qquad X := \begin{pmatrix} x \\ y \end{pmatrix} \qquad V := \begin{pmatrix} 0 \\ 27 \end{pmatrix} .$$

Note that

$$\vec{x} = \begin{pmatrix} \vec{e}_1 & \vec{e}_2 \end{pmatrix} X\,.$$

Also lets call

$$\vec{v} := \begin{pmatrix} \vec{e}_1 & \vec{e}_2 \end{pmatrix} V\,.$$

Now the matrix M is the matrix of some linear transformation L in the basis of the engineers. Lets convert it to the basis of the nutritionists:

$$L\vec{x} = L\begin{pmatrix} \vec{f}_1 & \vec{f}_2 \end{pmatrix}\begin{pmatrix} s \\ f \end{pmatrix} = L\begin{pmatrix} \vec{e}_1 & \vec{e}_2 \end{pmatrix} P \begin{pmatrix} s \\ f \end{pmatrix} = \begin{pmatrix} \vec{e}_1 \\ \vec{e}_2 \end{pmatrix} MP \begin{pmatrix} s \\ f \end{pmatrix}\,.$$

Note here that the linear transformation on acts on *vectors* -- these are the objects we have written with a $\vec{\ }$ sign on top of them. It does not act on columns of numbers!

We can easily compute MP and find

$$MP = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} -\frac{1}{\lambda} & 2 \\ \frac{1}{\lambda} & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{3}{\lambda} & 5 \end{pmatrix}\,.$$

Note that $P^{-1}MP$ is the matrix of L in the nutritionists basis, but we don't need this quantity right now.

Thus the last task is to solve the system, lets solve for sugar and fruit. We need to solve

$$MP\begin{pmatrix} s \\ f \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{3}{\lambda} & 5 \end{pmatrix}\begin{pmatrix} s \\ f \end{pmatrix} = \begin{pmatrix} 27 \\ 0 \end{pmatrix}\,.$$

This is solved immediately by forward substitution (the nutritionists basis is nice since it directly gives f):

$$f = 27 \quad \text{and} \quad s = 45\lambda\,.$$

2×2 Example

Lets diagonalize the matrix M from a previous example



Eigenvalues and Eigenvectors: 2×2 Example



$$M = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$$

We found the eigenvalues and eigenvectors of M, our solution was

$$\lambda_1 = 5,\ \mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \ \text{and}\ \lambda_2 = 2,\ \mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

So we can diagonalize this matrix using the formula $D = P^{-1}MP$ where $P = (\mathbf{v}_1, \mathbf{v}_2)$. This means

$$P = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \text{ and } P^{-1} = -\frac{1}{3}\begin{pmatrix} -1 & -1 \\ -1 & 2 \end{pmatrix}$$

The inverse comes from the formula for inverses of 2×2 matrices:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \text{ so long as } ad - bc \neq 0.$$

So we get:

$$D = -\frac{1}{3}\begin{pmatrix} -1 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}$$

But this does not really give any intuition into why this happens. Let look at what happens when we apply this matrix $D = P^{-1}MP$ to a vector $v = \begin{pmatrix} x \\ y \end{pmatrix}$. Notice that applying P translates $v = \begin{pmatrix} x \\ y \end{pmatrix}$ into $x\mathbf{v}_1 + y\mathbf{v}_2$.

$$\begin{aligned} P^{-1}MP\begin{pmatrix} x \\ y \end{pmatrix} &= P^{-1}M\begin{pmatrix} 2x + y \\ x - y \end{pmatrix} \\ &= P^{-1}M\left[\begin{pmatrix} 2x \\ x \end{pmatrix} + \begin{pmatrix} y \\ -y \end{pmatrix}\right] \\ &= P^{-1}\left[x\,M\begin{pmatrix} 2 \\ 1 \end{pmatrix} + y\,M\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right] \\ &= P^{-1}\left[x\,M\mathbf{v}_1 + y\,M\mathbf{v}_2\right] \end{aligned}$$

Remember that we know what M does to $\mathbf{v}_1$ and $\mathbf{v}_2$, so we get

$$\begin{aligned} P^{-1}[x\,M\mathbf{v}_1 + y\,M\mathbf{v}_2] &= P^{-1}[x\lambda_1\,\mathbf{v}_1 + y\lambda_2\,\mathbf{v}_2] \\ &= 5x\,P^{-1}\mathbf{v}_1 + 2y\,P^{-1}\mathbf{v}_2 \\ &= 5x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2y\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 5x \\ 2y \end{pmatrix} \end{aligned}$$

Notice that multiplying by P^{-1} converts $\mathbf{v}_1$ and $\mathbf{v}_2$ back in to $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ respectively. This shows us why $D = P^{-1}MP$ should be the diagonal matrix:

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix}$$

G.13 Orthonormal Bases and Complements

All Orthonormal Bases for $\mathbb{R}^2$

We wish to find all orthonormal bases for the space $\mathbb{R}^2$, and they are $\{e_1^\theta, e_2^\theta\}$ up to reordering where

$$e_1^\theta = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad e_2^\theta = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix},$$

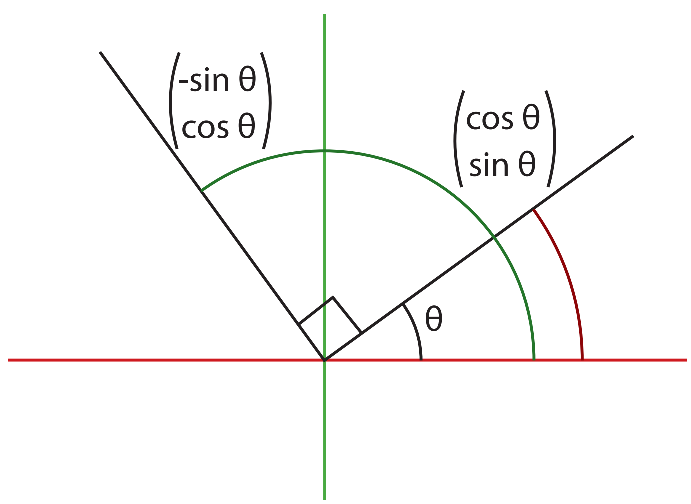
for some $\theta \in [0, 2\pi)$. Now first we need to show that for a fixed θ that the pair is orthogonal:

$$e_1^\theta \cdot e_2^\theta = -\sin\theta\cos\theta + \cos\theta\sin\theta = 0.$$

Also we have

$$\|e_1^\theta\|^2 = \|e_2^\theta\|^2 = \sin^2\theta + \cos^2\theta = 1,$$

and hence $\{e_1^\theta, e_2^\theta\}$ is an orthonormal basis. To show that every orthonormal basis of $\mathbb{R}^2$ is $\{e_1^\theta, e_2^\theta\}$ for some θ, consider an orthonormal basis $\{b_1, b_2\}$ and note that b_1 forms an angle ϕ with the vector e_1 (which is e_1^0). Thus $b_1 = e_1^\phi$ and if $b_2 = e_2^\phi$, we are done, otherwise $b_2 = -e_2^\phi$ and it is the reflected version. However we can do the same thing except starting with b_2 and get $b_2 = e_1^\psi$ and $b_1 = e_2^\psi$ since we have just interchanged two basis vectors which corresponds to a reflection which picks up a minus sign as in the determinant.



A 4×4 Gram Schmidt Example

Lets do an example of how to "Gram-Schmidt" some vectors in $\mathbb{R}^4$. Given the following vectors

$$v_1 = \begin{pmatrix} o \\ 1 \\ 0 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 3 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \text{ and } v_4 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix},$$

we start with v_1

$$v_1^\perp = v_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Now the work begins

$$\begin{aligned} v_2^\perp &= v_2 - \frac{(v_1^\perp \cdot v_2)}{\|v_1^\perp\|^2} v_1^\perp \\ &= \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{1} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \end{aligned}$$

This gets a little longer with every step.

$$\begin{aligned} v_3^\perp &= v_3 - \frac{(v_1^\perp \cdot v_3)}{\|v_1^\perp\|^2} v_1^\perp - \frac{(v_2^\perp \cdot v_3)}{\|v_2^\perp\|^2} v_2^\perp \\ &= \begin{pmatrix} 3 \\ 0 \\ 1 \\ 0 \end{pmatrix} - \frac{0}{1} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{1} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

This last step requires subtracting off the term of the form $\frac{u \cdot v}{u \cdot u}\mathbf{u}$ for each of the previously defined basis vectors.

$$\begin{aligned} v_4^\perp &= v_4 - \frac{(v_1^\perp \cdot v_4)}{\|v_1^\perp\|^2} v_1^\perp - \frac{(v_2^\perp \cdot v_4)}{\|v_2^\perp\|^2} v_2^\perp - \frac{(v_3^\perp \cdot v_4)}{\|v_3^\perp\|^2} v_3^\perp \\ &= \begin{pmatrix}1\\1\\0\\2\end{pmatrix} - \frac{1}{1}\begin{pmatrix}0\\1\\0\\0\end{pmatrix} - \frac{0}{1}\begin{pmatrix}0\\0\\1\\0\end{pmatrix} - \frac{3}{9}\begin{pmatrix}3\\0\\0\\0\end{pmatrix} \\ &= \begin{pmatrix}0\\0\\0\\2\end{pmatrix} \end{aligned}$$

Now $v_1^\perp$, $v_2^\perp$, $v_3^\perp$, and $v_4^\perp$ are an orthogonal basis. Notice that even with very, very nice looking vectors we end up having to do quite a bit of arithmetic. This a good reason to use programs like matlab to check your work.

Another QR Decomposition Example

We can alternatively think of the QR decomposition as performing the Gram-Schmidt procedure on the *column space*, the vector space of the column vectors of the matrix, of the matrix M. The resulting orthonormal basis will be stored in Q and the negative of the coefficients will be recorded in R. Note that R is upper triangular by how Gram-Schmidt works. Here we will explicitly do an example with the matrix

$$M = \begin{pmatrix} | & | & | \\ m_1 & m_2 & m_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ -1 & 1 & 1 \end{pmatrix}.$$

First we normalize m_1 to get $m_1' = \frac{m_1}{\|m_1\|}$ where $\|m_1\| = r_1^1 = \sqrt{2}$ which gives the decomposition

$$Q_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & 1 & -1 \\ 0 & 1 & 2 \\ -\frac{1}{\sqrt{2}} & 1 & 1 \end{pmatrix}, \qquad R_1 = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Next we find

$$t_2 = m_2 - (m_1' \bullet m_2) m_1' = m_2 - r_2^1 m_1' = m_2 - 0 m_1'$$

noting that

$$m_1' \bullet m_1' = \|m_1'\|^2 = 1$$

and $\|t_2\| = r_2^2 = \sqrt{3}$, and so we get $m_2' = \frac{t_2}{\|t_2\|}$ with the decomposition

$$Q_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -1 \\ 0 & \frac{1}{\sqrt{3}} & 2 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & 1 \end{pmatrix}, \qquad R_2 = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally we calculate

$$t_3 = m_3 - (m_1' \cdot m_3)m_1' - (m_2' \cdot m_3)m_2'$$
$$= m_3 - r_3^1 m_1' - r_3^2 m_2' = m_3 + \sqrt{2}m_1' - \frac{2}{\sqrt{3}}m_2',$$

again noting $m_2' \cdot m_2' = \|m_2'\| = 1$, and let $m_3' = \frac{t_3}{\|t_3\|}$ where $\|t_3\| = r_3^3 = 2\sqrt{\frac{2}{3}}$. Thus we get our final $M = QR$ decomposition as

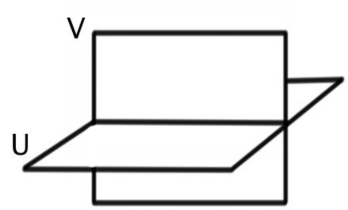
$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{3}} & \sqrt{\frac{2}{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{3} & -\frac{1}{\sqrt{6}} \end{pmatrix}, \qquad R = \begin{pmatrix} \sqrt{2} & 0 & -\sqrt{2} \\ 0 & \sqrt{3} & \frac{2}{\sqrt{3}} \\ 0 & 0 & 2\sqrt{\frac{2}{3}} \end{pmatrix}.$$

Overview

This video depicts the ideas of a subspace sum, a direct sum and an orthogonal complement in $\mathbb{R}^3$. Firstly, lets start with the subspace sum. Remember that even if U and V are subspaces, their union $U \cup V$ is usually not a subspace. However, the span of their union certainly is and is called the subspace sum

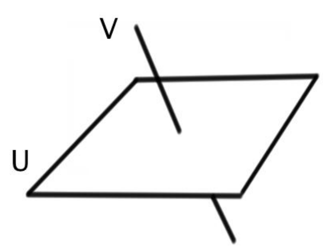
$$U + V = \text{span}(U \cup V).$$

You need to be aware that this is a sum of vector spaces (not vectors). A picture of this is a pair of planes in $\mathbb{R}^3$:



Here $U + V = \mathbb{R}^3$.

Next lets consider a direct sum. This is just the subspace sum for the case when $U \cap V = \{0\}$. For that we can keep the plane U but must replace V by a line:

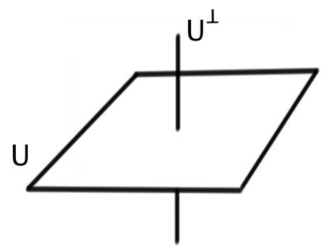


Taking a direct sum we again get the whole space, $U \oplus V = \mathbb{R}^3$.

Now we come to an orthogonal complement. There is not really a notion of subtraction for subspaces but the orthogonal complement comes close. Given U it provides a space $U^\perp$ such that the direct sum returns the whole space:

$$U \oplus U^\perp = \mathbb{R}^3 .$$

The orthogonal complement $U^\perp$ is the subspace made from all vectors perpendicular to any vector in U. Here, we need to just tilt the line V above until it hits U at a right angle:



Notice, we can apply the same operation to $U^\perp$ and just get U back again, *i.e.*

$$\left(U^\perp\right)^\perp = U .$$

Hint for Review Question 2

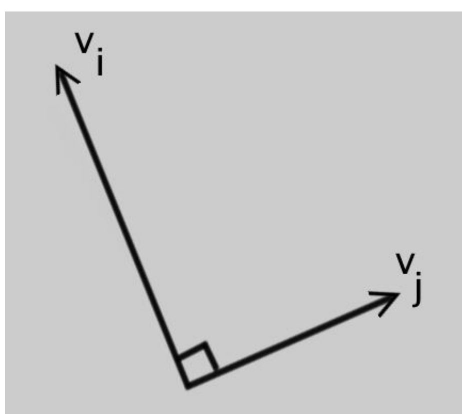
You are asked to consider an orthogonal basis $\{v_1, v_2, \ldots v_n\}$. Because this is a basis any $v \in V$ can be uniquely expressed as

$$v = c^1 v_1 + c^2 v_2 + \cdots + v^n c_n ,$$

and the number $n = \dim V$. Since this is an orthogonal basis

$$v_i \cdot v_j = 0 , \qquad i \neq j .$$

So different vectors in the basis are orthogonal:



However, the basis is *not* orthonormal so we know nothing about the lengths of the basis vectors (save that they cannot vanish).

To complete the hint, lets use the dot product to compute a formula for c^1 in terms of the basis vectors and v. Consider

$$v_1 \cdot v = c^1 v_1 \cdot v_1 + c^2 v_1 \cdot v^2 + \cdots + c^n v_1 \cdot v_n = c^1 v_1 \cdot v_1 \,.$$

Solving for c^1 (remembering that $v_1 \cdot v_1 \neq 0$) gives

$$c^1 = \frac{v_1 \cdot v}{v_1 \cdot v_1} \,.$$

This should get you started on this problem.

Hint for Review Problem 3

Lets work part by part:

(a) Is the vector $v^{\perp} = v - \frac{u \cdot v}{u \cdot u} u$ in the plane P?

Remember that the dot product gives you a scalar not a vector, so if you think about this formula $\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}$ is a scalar, so this is a linear combination of $\mathbf{v}$ and $\mathbf{u}$. Do you think it is in the span?

(b) What is the angle between $v^{\perp}$ and u?

This part will make more sense if you think back to the dot product formulas you probably first saw in multivariable calculus. Remember that

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\theta),$$

and in particular if they are perpendicular $\theta = \frac{\pi}{2}$ and $\cos(\frac{\pi}{2}) = 0$ you will get $\mathbf{u} \cdot \mathbf{v} = 0$.

Now try to compute the dot product of u and $v^{\perp}$ to find $\|\mathbf{u}\| \|\mathbf{v}^{\perp}\| \cos(\theta)$

$$\begin{aligned} u \cdot v^{\perp} &= u \cdot \left(v - \frac{u \cdot v}{u \cdot u} u \right) \\ &= u \cdot v - u \cdot \left(\frac{u \cdot v}{u \cdot u} \right) u \\ &= u \cdot v - \left(\frac{u \cdot v}{u \cdot u} \right) u \cdot u \end{aligned}$$

Now you finish simplifying and see if you can figure out what θ has to be.

(c) Given your solution to the above, how can you find a third vector perpendicular to both u and $v^{\perp}$?

Remember what other things you learned in multivariable calculus? This might be a good time to remind your self what the cross product does.

(d) Construct an orthonormal basis for $\mathbb{R}^3$ from u and v.

If you did part (c) you can probably find 3 orthogonal vectors to make a orthogonal basis. All you need to do to turn this into an orthonormal basis is make these into unit vectors.

(e) Test your abstract formulae starting with

$$u = \begin{pmatrix} 1 & 2 & 0 \end{pmatrix} \text{ and } v = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}.$$

Try it out, and if you get stuck try drawing a sketch of the vectors you have.

Hint for Review Problem 10

This video shows you a way to solve problem 10 that's different to the method described in the Lecture. The first thing is to think of

$$M = \begin{pmatrix} 1 & 0 & 2 \\ -1 & 2 & 0 \\ -1 & 2 & 2 \end{pmatrix}$$

as a set of 3 vectors

$$v_1 = \begin{pmatrix} 0 \\ -1 \\ -1 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 0 \\ 2 \\ -2 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}.$$

Then you need to remember that we are searching for a decomposition

$$M = QR$$

where Q is an orthogonal matrix. Thus the upper triangular matrix $R = Q^T M$ and $Q^T Q = I$. Moreover, orthogonal matrices perform rotations. To see this compare the inner product $u \cdot v = u^T v$ of vectors u and v with that of Qu and Qv:

$$(Qu) \cdot (Qv) = (Qu)^T (Qv) = u^T Q^T Q v = u^T v = u \cdot v.$$

Since the dot product doesn't change, we learn that Q does not change angles or lengths of vectors.

Now, here's an interesting procedure: rotate v_1, v_2 and v_3 such that v_1 is along the x-axis, v_2 is in the xy-plane. Then if you put these in a matrix you get something of the form

$$\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$$

which is exactly what we want for R!

Moreover, the vector

$$\begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix}$$

is the rotated v_1 so must have length $||v_1|| = \sqrt{3}$. Thus $a = \sqrt{3}$.

The rotated v_2 is

$$\begin{pmatrix} b \\ d \\ 0 \end{pmatrix}$$

and must have length $||v_2|| = 2\sqrt{2}$. Also the dot product between

$$\begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} b \\ d \\ 0 \end{pmatrix}$$

is ab and must equal $v_1 \cdot v_2 = 0$. (That v_1 and v_2 were orthogonal is just a coincidence here... .) Thus $b = 0$. So now we know most of the matrix R

$$R = \begin{pmatrix} \sqrt{3} & 0 & c \\ 0 & 2\sqrt{2} & e \\ 0 & 0 & f \end{pmatrix} .$$

You can work out the last column using the same ideas. Thus it only remains to compute Q from

$$Q = MR^{-1} .$$

G.14 Diagonalizing Symmetric Matrices

3×3 Example

Lets diagonalize the matrix

$$M = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

If we want to diagonalize this matrix, we should be happy to see that it is symmetric, since this means we will have real eigenvalues, which means factoring won't be too hard. As an added bonus if we have three distinct eigenvalues the eigenvectors we find will automatically be orthogonal, which means that the inverse of the matrix P will be easy to compute. We can start

by finding the eigenvalues of this

$$\begin{aligned}
\det\begin{pmatrix}1-\lambda & 2 & 0\\ 2 & 1-\lambda & 0\\ 0 & 0 & 5-\lambda\end{pmatrix} &= (1-\lambda)\begin{vmatrix}1-\lambda & 0\\ 0 & 5-\lambda\end{vmatrix}\\
&\quad -(2)\begin{vmatrix}2 & 0\\ 0 & 5-\lambda\end{vmatrix}+0\begin{vmatrix}2 & 1-\lambda\\ 0 & 0\end{vmatrix}\\
&= (1-\lambda)(1-\lambda)(5-\lambda)+(-2)(2)(5-\lambda)+0\\
&= (1-2\lambda+\lambda^2)(5-\lambda)+(-2)(2)(5-\lambda)\\
&= ((1-4)-2\lambda+\lambda^2)(5-\lambda)\\
&= (-3-2\lambda+\lambda^2)(5-\lambda)\\
&= (1+\lambda)(3-\lambda)(5-\lambda)
\end{aligned}$$

So we get $\lambda=-1,3,5$ as eigenvectors. First find v_1 for $\lambda_1=-1$

$$(M+I)\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}2&2&0\\2&2&0\\0&0&6\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix},$$

implies that $2x+2y=0$ and $6z=0$,which means any multiple of $v_1=\begin{pmatrix}1\\-1\\0\end{pmatrix}$ is an eigenvector with eigenvalue $\lambda_1=-1$. Now for v_2 with $\lambda_2=3$

$$(M-3I)\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}-2&2&0\\2&-2&0\\0&0&4\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix},$$

and we can find that that $v_2=\begin{pmatrix}1\\1\\0\end{pmatrix}$ would satisfy $-2x+2y=0$, $2x-2y=0$ and $4z=0$.

Now for v_3 with $\lambda_3=5$

$$(M-5I)\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}-4&2&0\\2&-4&0\\0&0&0\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix},$$

Now we want v_3 to satisfy $-4x+2y=0$ and $2x-4y=0$, which imply $x=y=0$, but since there are no restrictions on the z coordinate we have $v_3=\begin{pmatrix}0\\0\\1\end{pmatrix}$.

Notice that the eigenvectors form an orthogonal basis. We can create an orthonormal basis by rescaling to make them unit vectors. This will help us

because if $P = [v_1, v_2, v_3]$ is created from orthonormal vectors then $P^{-1} = P^T$, which means computing P^{-1} should be easy. So lets say

$$v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix},\ v_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix},\ \text{and } v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

so we get

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } P^{-1} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

So when we compute $D = P^{-1}MP$ we'll get

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 2 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

Hint for Review Problem 1

For part (a), we can consider any complex number z as being a vector in $\mathbb{R}^2$ where complex conjugation corresponds to the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Can you describe $z\bar{z}$ in terms of $\|z\|$? For part (b), think about what values $a \in \mathbb{R}$ can take if $a = -a$? Part (c), just compute it and look back at part (a).

For part (d), note that $x^\dagger x$ is just a number, so we can divide by it. Parts (e) and (f) follow right from definitions. For part (g), first notice that every row vector is the (unique) transpose of a column vector, and also think about why $(AA^T)^T = AA^T$ for any matrix A. Additionally you should see that $\overline{x^T} = x^\dagger$ and mention this. Finally for part (h), show that

$$\frac{x^\dagger M x}{x^\dagger x} = \overline{\left(\frac{x^\dagger M x}{x^\dagger x}\right)^T}$$

and reduce each side separately to get $\lambda = \overline{\lambda}$.

G.15 Kernel, Range, Nullity, Rank

Invertibility Conditions

Here I am going to discuss some of the conditions on the invertibility of a matrix stated in Theorem 16.3.1. Condition 1 states that $X = M^{-1}V$ uniquely, which is clearly equivalent to 4. Similarly, every square matrix M uniquely

corresponds to a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^n$, so condition 3 is equivalent to condition 1.

Condition 6 implies 4 by the adjoint construct the inverse, but the converse is not so obvious. For the converse (4 implying 6), we refer back the proofs in Chapter 18 and 19. Note that if $\det M = 0$, there exists an eigenvalue of M equal to 0, which implies M is not invertible. Thus condition 8 is equivalent to conditions 4, 5, 9, and 10.

The map M is injective if it does not have a null space by definition, however eigenvectors with eigenvalue 0 form a basis for the null space. Hence conditions 8 and 14 are equivalent, and 14, 15, and 16 are equivalent by the Dimension Formula (also known as the Rank-Nullity Theorem).

Now conditions 11, 12, and 13 are all equivalent by the definition of a basis. Finally if a matrix M is not row-equivalent to the identity matrix, then $\det M = 0$, so conditions 2 and 8 are equivalent.

Hint for Review Problem 2

Lets work through this problem.

Let $L\colon V \to W$ be a linear transformation. Show that $\ker L = \{0_V\}$ if and only if L is one-to-one:

1. First, suppose that $\ker L = \{0_V\}$. Show that L is one-to-one.

   Remember what one-one means, it means whenever $L(x) = L(y)$ we can be certain that $x = y$. While this might seem like a weird thing to require this statement really means that each vector in the range gets mapped to a unique vector in the range.

   We know we have the one-one property, but we also don't want to forget some of the more basic properties of linear transformations namely that they are linear, which means $L(ax + by) = aL(x) + bL(y)$ for scalars a and b.

   What if we rephrase the one-one property to say whenever $L(x) - L(y) = 0$ implies that $x - y = 0$? Can we connect that to the statement that $\ker L = \{0_V\}$? Remember that if $L(v) = 0$ then $v \in \ker L = \{0_V\}$.

2. Now, suppose that L is one-to-one. Show that $\ker L = \{0_V\}$. That is, show that 0_V is in $\ker L$, and then show that there are no other vectors in $\ker L$.

   What would happen if we had a nonzero kernel? If we had some vector v with $L(v) = 0$ and $v \neq 0$, we could try to show that this would contradict the given that L is one-one. If we found x and y with $L(x) = L(y)$, then we know $x = y$. But if $L(v) = 0$ then $L(x) + L(v) = L(y)$. Does this cause a problem?

G.16 Least Squares and Singular Values

Least Squares: Hint for Review Problem 1

Lets work through this problem. Let $L : U \to V$ be a linear transformation. Suppose $v \in L(U)$ and you have found a vector $u_{\rm ps}$ that obeys $L(u_{\rm ps}) = v$.

Explain why you need to compute $\ker L$ to describe the solution space of the linear system $L(u) = v$.

Remember the property of linearity that comes along with any linear transformation: $L(ax + by) = aL(x) + bL(y)$ for scalars a and b. This allows us to break apart and recombine terms inside the transformation.

Now suppose we have a solution x where $L(x) = v$. If we have an vector $y \in \ker L$ then we know $L(y) = 0$. If we add the equations together $L(x) + L(y) = L(x + y) = v + 0$ we get another solution for free. Now we have two solutions, is that all?

Hint for Review Problem 2

For the first part, what is the transpose of a 1×1 matrix? For the other two parts, note that $v \cdot v = v^T v$. Can you express this in terms of $\|v\|$? Also you need the trivial kernel only for the last part and just think about the null space of M. It might help to substitute $w = Mx$.

Index